UNIVERSIDAD NACIONAL DE COLOMBIA

Medición del perfil de riesgo de inversión usando un modelo de Teoría de Respuesta al Ítem bayesiano

David Hernando Medina Cifuentes

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C, Colombia
2020

Medición del perfil de riesgo de inversión usando un modelo de Teoría de Respuesta al Ìtem bayesiano

David Hernando Medina Cifuentes

Trabajo de grado presentado como requisito parcial para optar al título de:
Magíster en Ciencias Estadísticas

Director:
Ph.D. Alvaro Mauricio Montenegro Díaz

Línea de Investigación:
Teoría de respuesta al Ítem
Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C, Colombia
2020

Dedicatoria

Dedico a Dios mi trabajo.

Agradecimientos

Agradezco a Dios por guiarme en este sendero.

Resumen

En este trabajo se propone emplear un modelo en el marco de teoría de respuesta al ítem a través de estimación Bayesiana, con el objetivo de medir el trazo latente relacionado con el perfil de riesgo de inversión de las personas naturales que participan en el mercado financiero Colombiano. Para obtener los datos necesarios, se encuestaron a 144 individuos por medio de un cuestionario inicial compuesto por 18 ítems, construido a partir de una investigación rigurosa sobre los métodos y tipos de cuestionarios que emplean las principales entidades financieras del país para medir el perfil de riesgo de sus clientes. Se estimó un modelo *Generalized Partial Credit Model* (GPCM) en el lenguaje estadístico *Stan*, presentando buen desempeño en términos de convergencia de las cadenas de markov y de bondad de ajuste evaluado por medio del criterio predictivo p-valor bayesiano para los ítem finales y el modelo global.

Palabras clave: TRI,Bayes,Stan, Modelo Politómico.

Abstract

In this research we propose a model in the framework of item response theory through bayesian estimation, with the aim of measuring the latent trace related to the investment risk profile of natural persons participating in the Colombian financial market. For obtain the necessary data, 144 individuals were surveyed using an initial questionnaire made up of 18 items, constructed from rigorous research on the methods and types of questionnaires used by the country's main financial institutions to measure the risk profile of their clients. A Generalized Partial Credit Model was estimated in the statistical language *Stan*, showing good performance in terms of convergence of the markov chains and goodness of fit evaluated by means of the predictive bayesian p-value criterion for the finals items and the overall modell.

Keywords: TRI,Bayes,Stan,polytomous model.

Contenido

Agradecimientos iv

Resumen v

Lista de figuras vii

Lista de tablas 1

1. Introducción 2

2. Medición del perfil de riesgo 4
2.1. La pertinencia de medir el perfil de riesgo 4
2.2. ¿Cómo ha sido medido el perfil de riesgo? 5
2.3. Técnicas de medición en la academia y el sector financiero Colombiano . . . 8

3. Revisión de literatura TRI 10
3.1. Modelos de teoría de respuesta al ítem 11
3.1.1. Modelo de un parámetro (Rasch) 11
3.1.2. Modelo de dos parámetros 12
3.1.3. Modelo de tres parámetros 14
3.1.4. Modelos TRI politómicos 15
3.1.5. Modelos TRI multidimensionales 16
3.2. El enfoque bayesiano en los modelos TRI 17
3.2.1. El rol de la información a priori 18
3.2.2. Teorema de Bayes 18
3.2.3. Construyendo la posteriori 19
3.2.4. Modelo jerárquico bayesiano 21
3.2.5. Métodos MCMC 22
3.2.6. Estimación de los parámetros en los modelos TRI 25

4. Metodología 27
4.1. El instrumento y los datos 27
4.2. El modelo propuesto 28

5. Estimación y Resultados **30**
5.1. Algoritmo de estimación 31
5.2. Estimación de parámetros del modelo GPCM 32
5.3. Evaluación de ajuste del modelo 36
5.3.1. Diagnostico de convergencia de las cadenas de Markov 36
5.3.2. Ajuste del modelo 38

6. Conclusiones y recomendaciones **40**
6.1. Conclusiones 40

A. Anexo: Introducción al software estadístico bayesiano STAN **42**

B. Anexo: Constructo Inicial **46**

Bibliografía **50**

Lista de Figuras

3-1. Curvas características del ítem del modelo TRI de un parámetro con tres niveles de dificultad [Fox, 2010]. 12
3-2. Curvas características del ítem del modelo TRI de dos parámetros con tres niveles de discriminación y nivel igual de dificultad [Fox, 2010]. 13
3-3. Curvas características del ítem del modelo TRI de tres parámetro con tres niveles de adivinación y nivel igual de discriminación y dificultad [Fox, 2010]. 15

5-1. Análisis de componentes principales para el conjunto de ítems inicial. 30
5-2. Análisis de componentes principales para el conjunto de ítems inicial menos I,F,C y orden invertido para C. 31
5-3. Parámetros de dificultad estimados 32
5-4. CCCI ítem Q (14) 34
5-5. Bandas de credibilidad del 95 % para el parámetro θ_i 36
5-6. Correlación puntaje Clásico y θ_i estimados 37
5-7. Muestras de las cadenas estimando el trazo latente 37
5-8. Muestras de las cadenas estimando el trazo latente 38

Lista de Tablas

2-1. Teorías de medición clasificadas en dos tradiciones de investigación [Ruiz, 2016] 7

5-1. Estadísticas descriptivas de los parámetros de discriminación 33
5-2. Estadísticas descriptivas de los parámetros de dificultad 33
5-3. Estadística descriptiva del trazo latente 35
5-4. Clasificación de individuos según rango intercuartílico de las medias posteriores para el parámetro de habilidad. 36
5-5. Estadística descriptiva de Rhat del trazo latente 37
5-6. Resultados de ajuste (p valor bayesiano) para los ítems 39

1. Introducción

Debido al avance tecnológico, digital y crecimiento del sector financiero a nivel global, cada vez son más variadas las opciones y canales para que una persona natural pueda invertir sus recursos monetarios en diferentes tipologías de productos financieros, desde productos muy estables en rendimientos financieros como un certificado de deposito a término fijo (CDT) hasta acciones con altos rendimientos pero muy variables. Debido a esta variedad, muchas economías desarrolladas están empezando a exigir a las instituciones intermediarias[1], la realización de mediciones rigurosas y responsables para el perfil de riesgo de inversión de los clientes interesados en invertir sus ahorros en productos financieros como lo son CDT's, acciones y productos derivados entre otros, disponibles en cada bolsa de valores según el país de referencia.

Este crecimiento y avance del mercado financiero, ha impulsado la academia y a profesionales del sector en pro del desarrollo de medidas más sofisticadas desde el punto de vista cuantitativo y cualitativo. Estos esfuerzos están encaminados a entender y poder medir el comportamiento (perfil de riesgo) de los individuos según sus características económicas, psicológicas y de conocimiento financiero, bajo diferentes escenarios posibles de riesgo rentabilidad. Estas mediciones basadas en desarrollos rigurosos, permiten a las instituciones financieras poder perfilar y conocer mejor a sus clientes, y de esta forma formular mejores estrategias de inversión en términos de productos y montos acorde a las características propias (latentes) de cada cliente, lo cual protege al consumidor financiero y a su capital financiero.

En este trabajo propone una evaluación estadística para el perfil de riesgo de inversión de personas naturales, que sirva como instrumento de referencia para cumplir con la evaluación exigida por la superintendencia financiera de Colombia a los intermediarios entre el mercado de valores y los agentes inversionistas. Se consideran los aspectos más relevantes en la evaluación del perfil de riesgo como lo es el conocimiento financiero, la predisposición personal al riesgo-ganancia y el horizonte de tiempo planeado. A partir de la revisión del estado del

[1]La Bolsa de valores de Colombia define que los intervinientes en el mercado de valores comprenden fundamentalmente los emisores de valores, los intermediarios y los inversionistas. Dentro de estas relaciones que se dan entre emisores e inversionistas intervienen terceros que facilitan las transacciones entre los primeros, actuando así como intermediarios del mercado de valores, un ejemplo en Colombia puede ser un Banco o una entidad fiduciaria.

arte respecto a la evaluación del perfil de riesgo, se sugiere un modelo en el marco de la teoría de respuesta al ítem con estimación Bayesiana que supera en bondades al análisis clásico de test, además que por construcción permite obtener la capacidad de discriminación de los ítems y mediciones individuales del rasgo latente (perfil de riesgo de inversión) que sirve como insumo para representar la mejor posición financiera del individuo acorde a sus características psicológicas, económicas y personales.

2. Medición del perfil de riesgo

2.1. La pertinencia de medir el perfil de riesgo

En países como Estados Unidos, Italia y Holanda, las instituciones financieras ya están obligadas por la ley a construir perfiles de riesgo para sus clientes basandose en métodos cuantitativos formales, los cuales deben incluir aspectos del conocimiento financiero, horizonte de tiempo planeado y preferencias personales de riesgo [Caviezel et al., 2011]. Este perfilamiento se convierte en una herramienta que facilita a las instituciones realizar recomendaciones en las estrategias de inversión y tipos de productos adecuados para cada cliente dependiendo el riesgo dispuesto a asumir y características intrínsecas y económicas de cada cliente.

Para el caso del mercado de valores colombiano, si una persona natural decide realizar una inversión, esta no puede abordar directamente a un emisor de valores como la bolsa principal (BVC) y debe acudir a agentes intermediarios: comisionistas de bolsa, comisionistas independientes, establecimientos de crédito, las sociedades administradoras de fondos de inversión colectiva entre otros. Así como en las economías desarrolladas, en Colombia también se dispone de un marco legislativo (Decreto 1121 de 2008) que reglamenta toda actividad de intermediación entre instituciones y agentes inversionistas en el mercado de valores. En el artículo 1.5.3.3, se establece que será responsabilidad del intermediario establecer un perfil de riesgo del cliente y actuar de conformidad con el mismo, y cuando el intermediario considere que el producto o servicio ofrecido es inadecuado para el cliente, deberá darle a conocer expresamente su concepto, con el fin de proteger las inversiones de la persona.

Teniendo en cuenta el contexto presentado, la precisión y consistencia en las mediciones del perfil de riesgo de inversión, con variables extraídas de los clientes, se convierte en una prioridad no solo para las instituciones colombianas regidas por la ley, sino también para la academia y su investigación en el campo de la toma de decisiones de los individuos bajo escenarios de incertidumbre, específicamente en las decisiones que implican relaciones de riesgo-rentabilidad.

2.2. ¿Cómo ha sido medido el perfil de riesgo?

El estado del arte indica que la mayoría de los métodos de medición han sido cuantitativos, un grupo de técnicas teóricamente fundamentadas y otros aplicados empíricamente (e.g., scores, escalas e índices) [Ruiz, 2016]. Por otro lado, se han usado técnicas no teóricas y cualitativas que siguen vigentes como lo son evaluaciones subjetivas vía entrevistas por expertos financieros. Sin embargo, estas metodologías son cuestionables desde el punto de vista de la teoría de la medición del riesgo financiero.

La herramienta que más se ha empleado tanto en la academia como en la industria para la medición de la tolerancia al riesgo financiero han sido los cuestionarios de escala de calificación[1] [Ruiz, 2016]. Uno de los trabajos pioneros y más citados es el campo es el de Grabble and Lytton (1999), en cual se realiza una análisis de componentes principales a un grupo inicial de 20 ítems para finalmente quedarse con 13, los cuales logran medir tres aspectos en el individuo: 1) riesgo de inversión, 2) experiencia financiera y 3) riesgo especulativo, concluyendo en su trabajo que la herramienta ofrece un fuerte grado de dimensionalidad en la medición de la tolerancia al riesgo financiero. Un análisis comparativo del trabajo de Grabel y Lytton con el cuestionario desarrollado en la Universidad de Chicago se puede encontrar en Gilliam et al. (2010) en el cual se intenta medir el nivel de riesgo de los individuos a partir de un solo ítem con cuatro opciones de respuesta.

A partir del trabajo de Grabel Y Lytton, las herramientas desarrolladas para la medición de la tolerancia al riesgo financiero, se enfocaron en la generación de scores por medio de la teoría clasica del test (TCT) para posteriormente poder clasificar ordinalmente a los individuos en grupos de riesgo. Bajo la sombrilla de todas estas metodologías, se han establecido normalmente cuatro tipologías de preguntas para la medición: tipo de elecciones de inversión, una combinación de preguntas subjetivas y de inversión, evaluación del comportamiento real y bajo diferentes escenarios hipotéticos de riesgo-rentabilidad [Sherman et al., 2001]. Cabe mencionar que de forma paralela, se desarrollaron herramientas mixtas que emplean elementos tanto subjetivos como el criterio del experto financiero y elementos objetivos como cuestionarios orientados a los productos propios ofertados por las compañías financieras, sin embargo estas metodologías se encuentran dentro del marco de técnicas cualitativas de evaluación.

Hasta este punto, el recorrido por la literatura ha sido abordado desde la teoría de medición del riesgo financiero, pero es importante tener en cuenta que el concepto "perfil de riesgo"ha sido definido y medido desde diferentes campos disciplinares. Desde la rama de las finanzas

[1] Estos cuestionarios están compuestos por diferentes preguntas que evalúan la elección del individuo bajo ciertos escenarios de riesgo rentabilidad y se tienen alrededor de tres o cuatro opciones disponibles de respuesta por cada pregunta

el concepto empleado ha sido el de tolerancia al riesgo financiero, desde la economía y específicamente desde la teoría del consumidor, se habla de maximizar funciones de utilidad a partir de la elección de diferentes tipos de bienes. El trabajo de Hanna et al.(2001) ilustra como se asocia el concepto de utilidad esperada y aversión al riesgo. Otro campo que ha realizado un gran aporte y ha sido crucial es la Psicometría, con la formulación por parte de Charles Spearman de la Teoría Clásica del Test (TCT) en 1904 formulando las bases de la medición de test.

Investigadores han evidenciado las falencias de usar indiscriminadamente las herramientas empíricas hasta aquí mencionadas, que aunque son funcionales y fácilmente implementables, carecen de soporte teórico en la construcción, escalabilidad y validación. Un ejemplo de esto es el estudio de Grabble (2010) en donde se demuestra que que la herramienta desarrollada por la Universidad de Chicago que consta de un solo ítem presenta un bajo desempeño desde los criterios de la teoría clásica del test. Es por esto que se hace imperativo emplear teoría de medida y análisis psicométrico para potenciar las fases de desarrollo y evaluación de los instrumentos generados en la medición de la tolerancia al riesgo financiero de los agentes inversionistas [Ruiz, 2016].

A pesar de que previo al desarrollo de la teoría TCT muchas herramientas carecían de sustento teórico en la fase de desarrollo y evaluación, varios trabajos han incorporado teoría de medición en el desarrollo de instrumentos de evaluación financiera de inversión. Un ejemplo de esto es el instrumento desarrollado por la empresa Finametrica que proviene de la encuesta de tolerancia al riesgo desarrollada por Michael Roszkowski (1992) , el instrumento emplea modelos tradicionales de score, específicamente teoría clásica de test (TCT) como un primer acercamiento a una medida del perfil de riesgo. Otro trabajo que incorpora elementos de la teoría de medición es el trabajo de Grabble y Lytton (1999) , este describe los pasos de creación de la escala y la evaluación, empleando como técnica el análisis de componentes principales con el objetivo de reducir dimensionalidad de variables, este acercamiento también se enmarca dentro de la teoría clásica de test.

El conjunto de herramientas con fundamento teórico que persiguen realizar análisis de datos para modelar el perfil de riesgo en su mayoría se vinculan con la teoría clásica del test (TCT)[Ruiz, 2016], ya que una ventaja de la TCT es que se soporta en supuestos débiles y es relativamente fácil de interpretar. Sin embargo, TCT ha sido cuestionada ya que el verdadero score no es una característica absoluta del individuo evaluado y depende del contenido del test. Si hay individuos evaluados con diferente nivel de habilidad entonces el score final será diferente dependiendo del nivel de dificultad del test [Wiberg, 2004], otra crítica es que los parámetros de dificultad pueden variar dependiendo de la muestra de sujetos evaluados en un test especifico, esto quiere decir que si por casualidad se evalúan individuos con un nivel de habilidad muy alto, entonces los resultados del test indicaran que este era relativamente

Tabla 2-1.: Teorías de medición clasificadas en dos tradiciones de investigación [Ruiz, 2016]

Modelos de Score	**Modelos de escala**
Principales modelos: 1. Teoría clásica de test (TCT) 2. Teoría de la generalización 3. Análisis factorial (AC) 4. Modelos de ecuaciones estructurales (MEE)	Principales modelos: 1. Modelo psicofísico (MP) 2. Escala absoluta (EA) 3. Teoría de respuesta al ítem (TRI) 4. Teoría de respuesta al ítem no paramétrico
Principales propiedades: -Enfocados en el score del test -Modelos lineales -Enfocados en la estimación de los errores de los componentes	Principales propiedades: -Enfoque de respuesta ítem-persona -Modelos no lineales -Enfocados en modelar las respuestas de los individuos a los items
Autores clave: -Spearman (CTT) -Kuder and Richardson (CTT) -Cronbach and his colleagues (GT) -Spearman (FA) -Thurstone (FA) -Jöreskog (SEM)	Autores clave: -Thorndike (PM) -Thurstone (AS) -Birnbaum (IRT) -Rasch (IRT) -Guttman (NIRT) -Lazarsfeld (NIRT) -Mokken (NIRT)
Teoría en práctica: Brennan (GT) Jöreskog (SEM)	Teoría en práctica: Lord (IRT: Birnbaum) Wright (IRT: Rasch)

sencillo, esto por lo tanto, hace difícil la tarea de comparar los individuos entre diferentes test.

Con el objetivo de superar las falencias de la TCT, nuevos paradigmas se han desarrollado en las últimas cinco décadas. En el trabajo de Ruiz (2016) se menciona el paradigma de la tradición de escala como la teoría moderna de medición, en donde los modelos principales son la teoría de respuesta al ítem (TRI) y el modelo de medida de Rasch. La tabla $\mathbf{2-1}$ presenta un esquema detallado de los modelos empleados tanto en TCT como en el paradigma de escala.

TRI tiene varias ventajas sobre TCT, mientras las estadísticas de TCT dependen fundamentalmente del subconjunto de ítems y personas examinadas, los ítems en TRI y parámetros de las personas son invariantes, de esta forma se permite al investigador evaluar la contribución de ítems individuales en la medida que son añadidos u omitidos de un test. Adicionalmente, el modelo TRI puede distinguir el sesgo de los ítems de las diferencias reales sobre el atributo o aptitud medida, mientras que el modelo de TCT no puede realizar esto. Esto hace posible conducir test rigurosos de medidas equivalentes a través de grupos experimentales, lo cual es particularmente útil al realizar investigaciones en entornos interculturales donde diferencias de medias sobre el atributo son esperadas [Ostini and Nering, 2006]. En el capítulo de revisión teórica se expondrán en detalle las propiedades de los modelos bajo el marco de TRI.

A partir de los desarrollos teóricos de la teoría moderna de medición, en el marco de teoría de respuesta al ítem los campos que más han realizado uso extensivo de modelos son la psicología y las ciencias sociales, especialmente el sector educativo para el fortalecimiento de instrumentos aplicados y medición de la efectividad de los procesos pedagógicos. En el campo de evaluación financiera, muy pocos trabajos han tomado TRI para la construcción de instrumentos de medida, especialmente instrumentos con en el objeto de este trabajo que es la medición del perfil de riesgo de los inversionistas. Sin embargo, los trabajos de Caviezel et al. (2011) y de Ruiz (2006) proponen la construcción de medidas para el perfil de riesgo tomando elementos del modelo Partial credit model de Rasch.

2.3. Técnicas de medición en la academia y el sector financiero Colombiano

En Colombia la academia y la industria han empleado el marco de teoría de respuesta al ítem principalmente en aplicaciones orientadas a solucionar problemas en los campos de la psicometría, educación, salud, administración empresarial, mercadeo y economía. En el sector empresarial, el uso se ha enfocado en la evaluación de la percepción de la calidad de los servicios y el clima laboral. Sin embargo, muy pocos trabajos han realizado una formulación teórica para el área de evaluación financiera y específicamente en la evaluación del perfil de riesgo no se encuentran trabajos propuestos bajo el marco de TRI. Uno de los pocos trabajos formales que han propuesto un instrumento bajo la sombrilla de TRI en el área de evaluación financiera es la tesis de Rangel (2019), en donde se presenta un modelo politómico[2] unidimensional con el objeto de obtener medidas formales acerca del capital financiero de 384 microempresarios del área Metropolitana de Bucaramanga. Otra propuesta desarrollada desde la academia y que intenta medir el perfil de riesgo es el instrumento llamado *Conozca su perfil de riesgo* publicado por la editorial de la Unidad de Análisis del Mercado Financiero (UAMF) de la Facultad de Ciencias Económicas de la Universidad Nacional, en donde

[2] Los modelos politómicos fueron construidos para los ítems con respuesta politómica ordinales o nominales. Este tipo de respuestas proporcionan mayor información del trazo que las respuestas de tipo dicotómico. [Rangel, 2019]

se propone un sistema de ponderación subjetivo y clasificación ordinal de los inversionistas en tres categorías basándose en un score lineal que recoge las respuestas a siete ítems, esta propuesta no considera la TCT ni TRI como fundamento teórico.

Las empresas del sector financiero también presentan un escaso desarrollo en la fundamentación teórica de los instrumentos que emplean para evaluar el perfil de riesgo de sus inversionistas. En la investigación se recolectaron y analizaron ocho instrumentos de medida aplicados por cinco de los principales bancos en el mercado (Bancolombia, Davivienda, Scotiabank Colpatria e Itau) y tres sociedades administradoras de inversión autorizadas por la superintendencia financiera (Progresión S.A, AdCap Colombia S.A. Comisionista de Bolsa y Fiducoomeva). El análisis reveló una estructura y metodología de calculo similar en todos los instrumentos; la cual consiste en un score lineal que se calcula sumando los resultados individuales de cada ítem multiplicados por un factor de ponderación, que cambia en cada entidad acorde al criterio subjetivo o importancia que le otorguen al ítem evaluado. Posteriormente, de la misma forma que la herramienta propuesta por en la UAMF, se clasifica a los individuos en categorías ordinales según el score final y los puntos de corte definidos sin una metodología estadística o formal.

Si bien es cierto que las entidades financieras tienen la libertad de establecer sus límites de riesgo en el mercado de valores, la legislación en el decreto 1121 de 2008 establece que las entidades deben crear un perfil de riesgo objetivo para sus clientes y notificar cuando su perfil no sea compatible con el producto elegido por el cliente. Los resultados de la investigación anterior revelan que la metodología aplicada por las entidades financieras en primera medida rompen el principio de objetividad establecido por la ley al realizar una ponderación subjetiva de los ítems, emplear distinta cantidad de ítems sin un criterio formal y definir diferentes puntos de corte para la clasificación de los clientes. Además se evidencia la falta de investigación y rigurosidad por parte tanto del regulador como de las entidades en la construcción y mantenimiento de este tipo de herramientas tan cruciales para el desarrollo económico de las personas.

3. Revisión de literatura TRI

En la sociedad moderna, los test son usados extensivamente en el sector educativo, la industria y el gobierno. Los test tienen una variedad de funciones, pero a menudo se clasifican en dos grandes ramas: cognitivos (medidas de habilidad) y afectivos (diseñados para medir interés, actitudes y otros aspectos no cognitivos). En psicología, una de las formalizaciones más tempranas de una teoría de medida para test mentales es un conjunto de modelos matemáticos conocidos como teoría clásica del test (TCT) [Ostini and Nering, 2006]. TCT se deriva del trabajo pionero de Spearman (1904) que emplea elementos de medida de las ciencias físicas, ya que un concepto central de TCT es asumir la presencia de errores en el proceso de medición. Es por esto que los puntajes observados x_i son registrados con un error de medida e_i acorde a TCT como se observa en la ecuación 3-1:

$$\chi_i = \tau_i + e_i \tag{3-1}$$

En donde τ_i es el score verdadero (rasgo latente) que no puede ser observado directamente. Esta forma de establecer la relación del rasgo latente y las realizaciones de los ítems presentan falencias adicionales a las ya mencionadas en la sección 2.2 de este trabajo. El trabajo de Banchman (2004) describe cinco falencias centrales de la TCT, en donde la principal consiste en la dependencia del constructo y la muestra de individuos, lo que hace inviable conclusiones generalizadas fuera de muestra [Janseen et al., 2014]. Como resultado, las dificultades de validar los supuestos de la TCT y aplicar el modelo resultante en la práctica llevó al desarrollo de modelos de medición alternativos [Ostini and Nering, 2006].

El nacimiento de los modelos de respuesta al ítem en algunas partes de la literatura se presentan como una respuesta a las deficiencias de la teoría clásica de test [Hambleton et al., 1991]. La base formal de de TRI como una teoría del test orientada al ítem es atribuida normalmente al trabajo Lawley (1943) , cuyo trabajo fue expandido significativamente por Lord (1952) , quien también formalizó el papel de TRI como una extensión de la teoría clásica de test (TCT). El matemático danés Rasch jugó un papel igualmente influyente al desarrollar por separado una clase específica de modelos TRI y demostrar que tenía una serie de características altamente deseables [Ostini and Nering, 2006], estas características así como los principales modelos de TRI se explican en detalle en la siguiente sección.

3.1. Modelos de teoría de respuesta al ítem

Los modelos de TRI son concebidos con la medida de una variable latente hipotética (habilidad, capacidad o rasgo latente) que únicamente puede ser medida a través de otras variables manifiestas. Esta variable será descrita como θ en este trabajo. Los modelos de TRI tienen características deseables, la gran parte de estas características se derivan del hecho que una escala común es definida para la variable latente. Las características de los ítem y los individuos son parametrizadas de forma independiente además de ser invariantes[1], también la estimación de las características de ítems en diferentes muestras de individuos que pertenezcan a la misma población son comparables y difieren únicamente debido al error muestral [Fox, 2010].

Hay dos supuestos clave en TRI. El primero establece que el cambio en la variable latente que conlleva a un cambio en la probabilidad de una respuesta especifica, es descrito completamente por la curva característica del ítem[2] (CCI). El segundo supuesto establece que las respuestas a un par de ítems son estadísticamente independientes si la variable latente se mantiene constante. El supuesto de independencia local se establece en la ecuación 3-2:

$$P(\boldsymbol{Y_i}|\theta_i) = P(y_{i1}|\theta_i)P(y_{i2}|\theta_i)\dots P(y_{iK}|\theta_i) = \prod_{k=1}^{K} P(y_{ik}|\theta_i) \tag{3-2}$$

En donde se tiene un vector aleatorio de K respuestas denotado como $\boldsymbol{Y_i}$, con valores observados $c = (y_{i1}, y_{i2}, ..., y_{iK})$ de un individuo indexado i con parámetro de habilidad θ_i. De esta forma se establece la existencia de una variable latente y después de condicionar sobre esta, las respuestas observadas se asumen independientes. Por lo tanto, el supuesto de independencia local es también conocido como independencia condicional.

3.1.1. Modelo de un parámetro (Rasch)

El modelo de respuesta logística de un parámetro, también conocido en la literatura como el modelo de Rasch (Rasch, 1960), es uno de de lo modelos más simples y ampliamente usados en TRI, en este modelo la probabilidad de una respuesta correcta es dada por:

$$P(Y_{ik} = 1 \mid \theta_i, b_k) = \frac{exp(\theta_i - b_k)}{1 + exp(\theta_i - b_k)} = (1 + exp(b_k - \theta_i))^{-1} \tag{3-3}$$

Para un individuo i con nivel de habilidad θ_i y parámetro de dificultad del ítem b_k. En la

[1] Esto significa que las estimaciones son no test dependientes.

[2] Esta describe como la probabilidad de respuesta a un ítem cambia debido a variaciones en la variable latente.

gráfica **3-1** se pueden observar tres CCIs[3] correspondientes a la ecuación 3-3 con diferentes niveles de dificultad. Una característica importante de las CCIs en este modelo es que son paralelas, lo cual indica un nivel discriminación igual. Esto significa que en las tres curvas un aumento en el nivel de habilidad equivale a un aumento igual en la probabilidad de respuesta correcta.

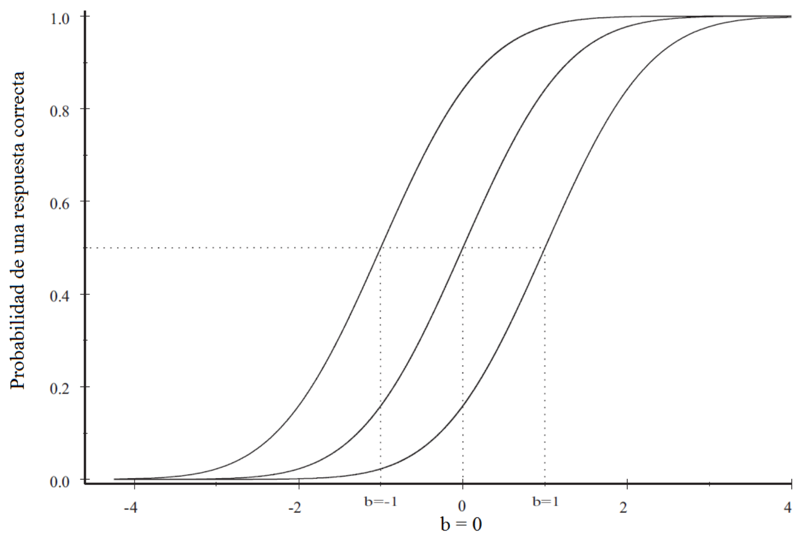

Figura 3-1.: Curvas características del ítem del modelo TRI de un parámetro con tres niveles de dificultad [Fox, 2010].

El parámetro de dificultad b_k es el punto en la escala de habilidad que corresponde a una probabilidad de respuesta correcta de 1/2. Se dice que un ítem es más fácil cuando la probabilidad de éxito es más alta en comparación con otro ítem dado el mismo nivel de habilidad.

El modelo de la ecuación 3-3 tiene unas características deseables, esto se debe a que la distribución de probabilidad hace parte de la familia exponencial de distribuciones. Como resultado, la estructura del modelo de Rasch permite la separación algebraica de la habilidad y los parámetros de los ítems. Por lo tanto, en la estimación de parámetros de los ítems, los parámetros de habilidad pueden ser eliminados a través de la estimación de máxima verosimilitud condicional.

3.1.2. Modelo de dos parámetros

Una limitación del modelo de un parámetro es que todos los ítem tienen el mismo nivel de discriminación. En el modelo logístico de dos parámetros, un parámetro de discriminación es añadido al modelo como se ve en la ecuación 3-4:

[3] Las curvas características del ítem (CCIs) describen la relación entre el nivel de habilidad y la probabilidad de una respuesta correcta.

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k) = \frac{exp(a_k\theta_i - b_k)}{1 + exp(a_k\theta_i - b_k)} = (1 + exp(b_k - a_k\theta_i))^{-1} \tag{3-4}$$

Como resultado, la curva característica del ítem (CCI) tiene parámetro de pendiente a_k y los ítem dejan de estar igualmente relacionados al parámetro de habilidad. En la figura **3-2**, tres curvas CCIs son graficadas con el mismo nivel de dificultad ($b_k = 0$) y tres valores de discrimiancón diferentes de 2, 1 y 1/2. La curva con nivel de discriminación 2 corresponde a la más empinada, lo cual evidencia que a un mayor nivel de discriminación mejor es la capacidad del ítem para detectar bajos niveles de habilidad versus altos niveles.

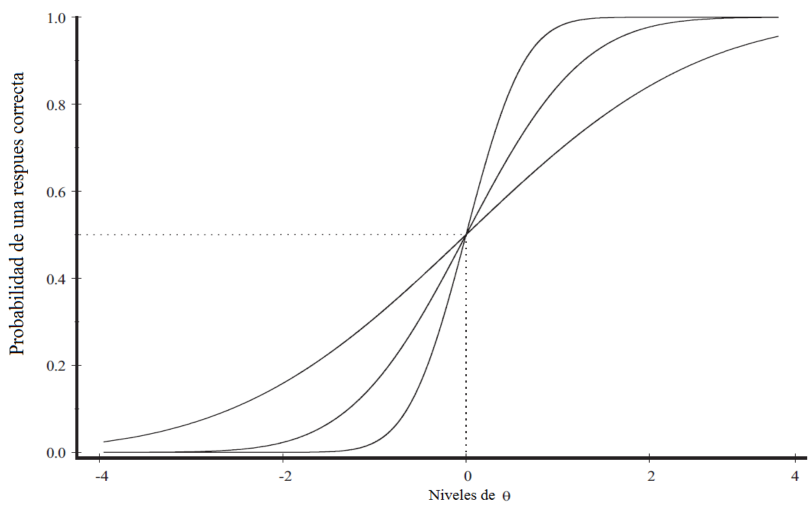

Figura 3-2.: Curvas características del ítem del modelo TRI de dos parámetros con tres niveles de discriminación y nivel igual de dificultad [Fox, 2010].

Para los parámetros de habilidad, en este modelo no hay estadísticos suficientes, por lo tanto la estimación de máxima verosimilitud condicional no es viable. Block y Liberman (1970) desarrollaron un proceso de estimación basado en máxima verosimilitud marginal (MML) para el modelo de dos parámetros. Los parámetros de los ítem son estimados de la distribución marginal, primero integrando sobre la distribución de habilidad y luego eliminando los parámetros de habilidad de la función de verosimilitud.

Una versión probit del modelo de dos parámetros es definido en la literatura como el modelo de ojiva normal, en el cual la CCI está basada en una distribución normal acumulada como se ilustra en la ecuación 3-5

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k) = \Phi(a_k\theta_i - b_k) = \int_{-\infty}^{a_k\theta_i - b_k} \phi(z)dz \tag{3-5}$$

En donde $\Phi(.)$ y $\phi(.)$ son la función de distribución acumulada normal y la función de densidad normal[4], respectivamente. Sin embargo, las CCIs para los modelos de las ecuaciones 3-4 y 3-5 se asemejan bastante cuando se multiplica los parámetros de los ítem en del modelo logístico por un factor constante escalar $d = 1{,}7$, de forma que la probabilidad de respuesta en las dos curvas difiere en menos de 0.01 [Hambleton et al., 1991, p. 15]

3.1.3. Modelo de tres parámetros

El modelo de dos parámetros puede ser extendido para permitir respuestas dadas al azar, esto se hace introduciendo una asíntota inferior distinta de cero para la CCI como se observa en la formula 3-6 y 3-7,

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k, c_k) = c_k + (1 - c_k)\Phi(a_k\theta_i - b_k) \tag{3-6}$$

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k, c_k) = \Phi(a_k\theta_i - b_k) + c_k(1 - \Phi(a_k\theta_i - b_k)) \tag{3-7}$$

En donde c_k es conocido como el parámetro de respuesta al azar para el ítem k. La probabilidad de una respuesta correcta es dada por un parámetro de adivinación más un segundo término que representa la probabilidad de una respuesta correcta dependiendo en los valores del parámetro del ítem y el nivel de habilidad del individuo i. La versión logística de este modelo se observa en la formula 3-8,

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k, c_k) = c_k + \frac{1 - c_k}{1 + exp(b_k - a_k\theta_i)} \tag{3-8}$$

$$P(Y_{ik} = 1 \mid \theta_i, a_k, b_k, c_k) = \frac{1}{1 + exp(b_k - a_k\theta_i)} + \frac{c_k}{1 + exp(a_k\theta_i - b_k)} \tag{3-9}$$

Se observa que cuando $c_k = 0$ el modelo de tres parámetros es igual al modelo de dos parámetros, En la figura **3-3** se pueden ver tres CCIs con diferente nivel de adivinación; bajo (0.05), medio (0.1) y alto (0.2) pero con el mismo nivel de dificultad y discriminación. La asíntota más baja de cada CCI indica el nivel de adivinación de dicho ítem y corresponde a la probabilidad de responder de forma correcta cuando el individuo adivina la respuesta.

En la aplicación de TRI a encuestas, este tipo de modelo no presenta utilidad, debido a que no hay respuestas incorrectas. Por eso no se profundiza más en este tipo de modelo en este trabajo.

[4] Una variable aleatoria Z es normalmente distribuida con media μ y varianza σ^2 cuando su función de densidad de probabilidad es igual a $\phi(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}exp(\frac{-1}{2\pi\sigma^2}(z-\mu)^2)$. La función de densidad normal estándar es definida cuando μ=0 y σ=1

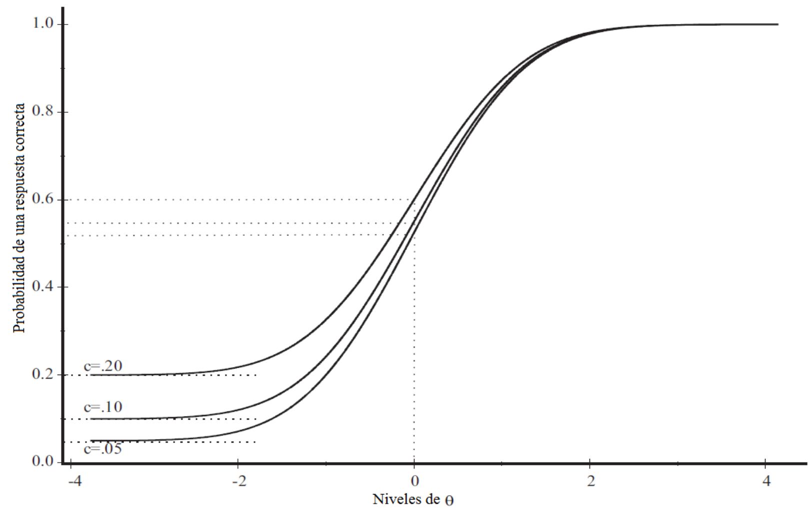

Figura 3-3.: Curvas características del ítem del modelo TRI de tres parámetro con tres niveles de adivinación y nivel igual de discriminación y dificultad [Fox, 2010].

3.1.4. Modelos TRI politómicos

Los modelos presentados hasta este punto son de forma binaria[5]. Sin embargo, a menudo los ítems son aplicados con múltiples categorías en donde cada categoría es punteada de forma separada o en donde existen intensidades de respuesta, que varían en un rango cualitativo. Los modelos politómicos se basan en su mayoría en ítems politómicos ordenados, que son ítems en donde la categorías de respuestas pueden ser ordenadas de acuerdo al nivel de habilidad[6]. Una ventaja de este tipo de modelos es que aunque se tienen más parámetros, se puede obtener información más precisa acerca del nivel de habilidad. El modelo más común para respuestas politómicas es el Partial Credit Model (PCM) [Masters, 1982]. La probabilidad de una respuesta en una categoría particular $c(c = 1, ..., C_k)$ del ítem k es definida directamente como

$$P(Y_{ik} = c \mid \theta_i, \boldsymbol{\kappa}_k) = \frac{exp \sum_{l=1}^{c} (\theta_i - \kappa_{k,l})}{\sum_{r=1}^{C_k} (exp \sum_{l=1}^{r} (\theta_i - \kappa_{k,l}))}, \tag{3-10}$$

en donde $\kappa_{k,l}$ es el parámetro de dificultad del paso del ítem y $\sum_{l=1}^{1} (\theta_i - \kappa_{k,l}) \equiv 0$. El número de categorías por ítem puede ser diferente. El modelo PCM se simplifica al modelo de un parámetro cuando las categorías son dos. En el modelo PCM, las probabilidades acumuladas no son modeladas directamente pero sí son el resultado de sumar las funciones de respuesta de las categorías. En el modelo de respuesta graduada, las probabilidades acumuladas son modeladas directamente.

[5]Estos son modelos que consideran ítems en donde se tiene una respuesta correcta o incorrecta

[6]Modelos de respuestas con ítems politómicos ordenados también son llamados modelos de respuesta graduada

En el modelo de respuesta graduada [Samemija, 1997] la probabilidad de responder cierta categoría es modelada por la probabilidad de responder por encima de la categoría anterior menos la probabilidad de responder por encima de la categoría. Sea C_k el número de categorías de respuesta en el ítem k. Entonces, hay $C_k - 1$ umbrales entre las opciones de respuestas. El modelo de respuesta graduada tiene la representación matemática en la expresión 3-11,

$$\begin{aligned} P(Y_{ik} = c \mid \theta_i, \boldsymbol{\kappa}_k) &= (Y_{ik} \geq c-1 \mid \theta_i, \boldsymbol{\kappa}_k) - (Y_{ik} \geq c \mid \theta_i, \boldsymbol{\kappa}_k) \\ &= \int_{\kappa_{k,c-1}}^{\infty} \psi(z; a_k\theta_i) dz - \int_{\kappa_{k,c}}^{\infty} \psi(z; a_k\theta_i) dz \\ &= \Psi(a_k\theta_i - \kappa_{k,c-1}) - \Psi(a_k\theta_i - \kappa_{k,c}) \\ &= \frac{exp(a_k\theta_i - \kappa_{k,c-1})}{1 + exp(a_k\theta_i - \kappa_{k,c-1})} - \frac{exp(a_k\theta_i - \kappa_{k,c})}{1 + exp(a_k\theta_i - \kappa_{k,c})}, \end{aligned} \tag{3-11}$$

en donde ψ y Ψ son las funciones de densidad[7] y distribución acumulada logística respectivamente. La probabilidad de tener un puntaje por arriba de la categoría más baja es uno y por encima de la más alta es cero. El modelo de respuesta graduada tiene una restricción de orden sobre los parámetros de los umbrales en comparación con el modelo Partial Credit Model.

3.1.5. Modelos TRI multidimensionales

Algunos constructos o test requieren de múltiples habilidades para obtener una respuesta correcta. Un buen ejemplo es una prueba matemática en donde sea necesario leer una historia para resolver el problema. En este tipo de prueba se evalúan habilidades tanto matemáticas como de compresión lectora. En primera medida, la probabilidad de obtener una respuesta correcta a un test no es decreciente cuando se elevan las habilidades de cada uno de los ítems. Esto se refiere al supuesto de monoticidad para los modelos unidimensionales. En segunda medida, las respuestas individuales a los ítems son condicionalmente independientes dados los valores de habilidad del individuo, lo cual es el el supuesto de independencia local.Considerando estos supuestos, la forma básica de un modelo TRI multidimensional para respuestas binarias es una generalización del modelo TRI unidimensional.

En esta generalización, cada individuo es descrito por múltiples parámetros más que por un parámetro escalar único. Esta extensión a múltiples dimensiones del modelo unidimensional

[7]Una variable Z está logísticamente distribuida con media μ y varianza $\sigma^2\pi^2/3$ cuando su función de densidad de probabilidad es igual $\psi(z; \mu, \sigma^2) = \frac{exp((z-\mu)/\sigma)}{\sigma(1+exp((z-\mu)/\sigma))^2}$. La función de densidad logística estándar es definida cuando $\mu = 0$ y $\sigma = 1$.

de dos parámetros tiene la representación matemática en la ecuación 3-12:

$$\begin{aligned} P(Y_{ik} = 1 \mid \boldsymbol{\theta}_i, \boldsymbol{a}_k, b_k) &= \frac{exp(\sum_q a_{kq}\theta_{iq} - b_k)}{1 + exp(\sum_q a_{kq}\theta_{iq} - b_k)} \\ &= \frac{exp(\boldsymbol{a}_k^t \theta_i - b_k)}{1 + exp(\boldsymbol{a}_k^t \theta_i - b_k)}, \end{aligned} \tag{3-12}$$

en donde cada individuo i tiene un vector de parámetros de habilidad θ_i con elementos $\theta_{i1}, ..., \theta_{iQ}$. Los elementos de la matriz de discriminación para el ítem k, $\boldsymbol{a}_k$, pueden ser interpretados como el poder de discriminación del ítem. El nivel de discriminación, a_{kq}, refleja el cambio en la probabilidad de una respuesta correcta debido a un cambio en la habilidad correspondiente θ_{iq}

3.2. El enfoque bayesiano en los modelos TRI

El fundamento de la estadística bayesiana recae el teorema de Thomas Bayes desarrollado en el siglo XVIII. Aunque este teorema es bastante útil, en el siglo XVIII y XIX tuvo un desarrollo más lento en comparación al siglo XX. Esto debido a los cálculos extensivos necesarios como la integración numérica. Entonces los avances más significativos después de Bayes fueron realizados por los trabajos de Laplace a inicios del siglo XIX.

En el siglo XX el desarrollo de las estadística bayesiana continuó con los trabajos de Jeffreys, Lindley, Box y Tiao. Sin embargo, en el momento de desarrollo de estos trabajos, los métodos de simulación por computadora eran más costosos y menos eficientes de lo que son ahora, por lo que restringieron su atención a las familias conjugadas y dedicaron mucho esfuerzo a derivar formas analíticas para las densidades posteriores marginales [Gao, 2018, pág 1].

Ahora en siglo XXI, gracias los avances computacionales y de software, se han superado las grandes barreras que existían en los siglos anteriores, y añadiendo el desarrollo de los métodos de muestreo intensivos, se ha revolucionado la estadística computacional y de la misma forma se ha fomentado la aplicación de los métodos bayesianos en diferentes áreas del conocimiento.

Con la instauración de estos postulados, desde el punto de vista teórico a diferencia del enfoque frecuentista, el enfoque bayesiano considera que los parámetros de los modelos son realizaciones de variables aleatorias y tienen distribuciones prior que reflejan la incertidumbre acerca de los parámetros. Lo modelos (TRI para este trabajo) propuestos para los datos observados describen el proceso de generación de datos como una función de parámetros desconocidos y son llamados modelos de probabilidad. Por lo tanto, dos etapas de modelamiento pueden ser reconocidas: (1) la especificación de una distribución prior y (2) la especificación de un modelo de probabilidad. Después de observar los datos, la información prior se combina con la información de los datos y una distribución posterior es construida[8].

3.2.1. El rol de la información a priori

Las distribuciones a priori de los parámetros desconocidos del modelo son especificadas de tal forma que capturan la creencia del investigador acerca de la situación antes de conocer los datos. La elección de información a priori puede ser debatible pero no es arbitraria ya que representa el conocimiento del investigador, teniendo en cuenta que también es posible incorporar distribuciones a priori que indiquen un desconocimiento total de los posibles valores de los parámetros (prior no informativas). Además, la información a priori puede basarse en los datos observados o nueva información relevante, lo que ayuda fortalecer la elección objetiva de la información a priori.

El uso de esta información a priori puede mejorar la confiabilidad de las inferencias estadísticas. Las respuestas se obtienen en un entorno real, y las fuentes de información externas a los datos se pueden incorporar a través de una distribución a priori . En los casos en donde se tenga poca información de bases de datos, la información a priori puede mejorar substancialmente las inferencias estadísticas [Fox, 2010].

Las distribuciones a priori elegidas dentro de los modelos bayesianos a menudo son clasificadas en tres categorías dependiendo la influencia ejercida en la distribución posterior y teniendo en cuanta su interacción con la verosimilitud de los datos. Un primer grupo son las distribuciones prior informativas que permiten resumir la información de los eventos pasados relacionados con θ. Un grupo que permite que jugar un rol mínimo en la distribución posterior son las prior no informativas, en donde el argumento principal es permitir a los datos hablar por ellos mismos [Gao, 2018], además que la distribución posterior coincide con la verosimilitud de los datos. Entre otros grupos usados están la distribuciones prior débiles las cuales se asemejan a las prior informativas pero su influencia es de menor impacto en la distribución posterior.

3.2.2. Teorema de Bayes

A través de un experimento estadístico los datos recolectados provienen de cada evento que tiene un resultado aleatorio o impredecible. Sean N observaciones denotadas por $\mathbf{y} = (y_1, ..., y_N)$, se asume que $\mathbf{y}$ es una realización numérica de un vector aleatorio $\mathbf{Y} = (Y_1, ..., Y_N)$ que tiene alguna distribución de probabilidad.[9] En este trabajo se asume que los

[8] Las inferencias bayesianas son realizadas condicionalmente sobre los datos y las inferencias acerca de los parámetros son realizadas directamente de las distribuciones posteriores

[9] Por simplicidad $\mathbf{Y}$ es un vector aleatorio continuo o discreto con función de probabilidad $p(\mathbf{y})$ para $\mathbf{y}$.

datos respuestas son usados para medir una variable latente $\boldsymbol{\theta}$ que representa características del individuo. La expresión $p(\boldsymbol{\theta})$ representa la información disponible a priori[10] sin conocer las realizaciones de los datos.

El término $p(\mathbf{y} \mid \boldsymbol{\theta})$ como función de los datos será la distribución muestral, y como función de parámetros será la función de verosimilitud. El interés en este caso se centra en la distribución de los parámetros $\boldsymbol{\theta}$ dados los datos observados. Esta distribución condicional de $\boldsymbol{\theta}$ dada la muestra se observa en la ecuación 3-13, este resultado matemático en teoría de probabilidad es conocido como el teorema de Bayes:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \tag{3-13}$$

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{3-14}$$

El término $p(\boldsymbol{\theta} \mid \mathbf{y})$ es la densidad posterior del parámetro $\boldsymbol{\theta}$ dada la creencia a priori y la información de la muestra. El término $p(\mathbf{y})$ corresponde a la densidad marginal de los datos y su evaluación computacionalmente es costosa. Cuando es suficiente conocer la forma de la posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$, la función de densidad posterior anormalizada puede ser usada como en la ecuación 3-14. Gracias a los métodos computacionales bayesianos, la densidad posterior de los parámetros $p(\boldsymbol{\theta} \mid \mathbf{y})$ es usada para realizar inferencias sin tener que apoyarse en aproximaciones asintóticas. Este enfoque bayesiano permite una forma de mejorarar la precisión de los parámetros estimados al combinar la información previa con la información de los datos (verosimilitud) [Fox, 2010].

3.2.3. Construyendo la posteriori

A modo de ilustración, se asume que se tienen cinco respuestas dicotómicas para cinco ítem $\mathbf{y} = (1, 1, 0, 0, 0)^t$ de un individuo con habilidad θ. El objetivo aquí es estimar la densidad posterior del parámetro de habilidad. Acorde a la versión probit del modelo de Rasch, con $P(Y_k = 1 \mid \theta) = \Phi(\theta)$ que define la probabilidad de una respuesta correcta para el ítem k. Se cree a priori que la probabilidad de dar una respuesta correcta es igual a la probabilidad de dar una respuesta incorrecta, por lo tanto θ tiene una distribución a priori uniformemente distribuida en el intervalo [-3,3] tal que $0{,}001 < \Phi(\theta) < 0{,}998$.

La función de verosimilitud para θ es igual $p(\mathbf{y} \mid \theta) = \Phi(\theta)^2(1 - \Phi(\theta))^3$ y multiplicando como en la expresión 3-14 la verosimilitud con la prior, la posterior corresponde la expresión 3-15:

$$p(y \mid \theta) = \Phi(\theta)^2(1 - \Phi(\theta))^3 \tag{3-15}$$

[10] Este término $p(\boldsymbol{\theta})$ es llamado la distribución a priori o simplemente la priori

Para $\theta \in [-3, 3]$. La moda a posterior, en la cual la densidad es maximizada, puede ser calculada en este caso tomando la primera derivada del logaritmo de la distribución posterior, igualando la expresión a cero, y resolviendo la ecuación final para θ.

El teorema de Bayes puede ser visto como una regla de actualización en donde los datos observados son usados para transformar las creencias a priori a creencias posteriores. Las observaciones son condicionalmente independientes dado θ y por lo tanto la posterior puede ser vista como en la expresión 3-16,

$$\begin{aligned} p(\theta \mid y_1, y_2 \ldots, y_k) &\propto p(y_1 \mid \theta)p(y_2 \mid \theta)...p(y_k \mid \theta)p(\theta) \\ &\propto p(\theta \mid y_1, y_2 \ldots, y_{k-1})p(y_k \mid \theta) \end{aligned} \tag{3-16}$$

La expresión anterior puede ser interpretada en sentido que la densidad posterior dada todas las observaciones menos la última, es actualizada a través de la verosimilitud de la última observación. En el marco de TRI, los parámetros de los individuos y de los ítems son el interés. El objetivo de las inferencias es la distribución posterior que a menudo es resumida reportando la media posterior y la desviación estándar. Entonces, en el caso de TRI por teorema de Bayes, la densidad conjunta posterior de los parámetros de interés puede ser vista como en la ecuación 3-17

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \mathbf{y}) &= p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta}, \boldsymbol{\xi})/p(\mathbf{y}) \\ &= p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta})p(\boldsymbol{\xi})/p(\mathbf{y}) \end{aligned} \tag{3-17}$$

En donde las densidades a priori se asumen independientes una de otra. Concluir estimaciones a partir densidad posterior conjunta es muy complicado debido a la alta dimensionalidad y a las formas analíticas difíciles de tratar. De esta forma cuando el interés recae en un parámetro como θ los parámetros de los ítem deben ser marginalizados como en la ecuación 3-18

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}) &= \int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta})p(\boldsymbol{\xi})/p(\mathbf{y})d\boldsymbol{\xi} \\ &= \int p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \mathbf{y})d\boldsymbol{\xi} \end{aligned} \tag{3-18}$$

Se necesitan evaluar más formas integrales cuando el interés recae en la densidad posterior marginal de un solo componente, como por ejemplo del vector de parámetros de individuos. Inclusive trabajar sobre las densidades marginales posteriores es una tarea difícil ya que las formas matemáticas no son conocidas. Es en este punto en donde métodos basados en simulación son capaces de generar muestras de las posteriores marginales, para posteriormente usar las muestras con fines de inferencia estadística.

3.2.4. Modelo jerárquico bayesiano

En un enfoque jerárquico bayesiano, los hiper parámetros de la densidad priori se caracterizan por poseer también una densidad a priori (hiper densidad) en una segunda etapa. Los hiper parametros son denotados por $\theta_p = (\mu_\theta, \sigma^2_\theta)$ y $\xi_p = (\mu_\xi, \Sigma_\xi)$. Entonces la densidad posterior de interés para los parámetros de las personas y los ítem puede ser expresada como en la ecuación 3-19

$$\begin{aligned} p(\boldsymbol{\xi}, \boldsymbol{\theta} \mid \mathbf{y}) &\propto \int\int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{\theta}_P, \boldsymbol{\xi}_P) p(\boldsymbol{\theta}_P, \boldsymbol{\xi}_P) d\boldsymbol{\xi}_P d\boldsymbol{\theta}_P \\ &\propto \int\int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\theta} \mid \boldsymbol{\theta}_P) p(\boldsymbol{\xi} \mid, \boldsymbol{\xi}_P) p(\boldsymbol{\theta}_P) p(\boldsymbol{\xi}_P) d\boldsymbol{\xi}_P d\boldsymbol{\theta}_P \\ &\propto \int\int \prod_{i,k} [p(y_{ik} \mid \theta_i, \xi_k) p(\theta_i \mid \boldsymbol{\theta}_P) p(\xi_k \mid \boldsymbol{\xi}_P)] p(\boldsymbol{\theta}_P) p(\boldsymbol{\xi}_P) d\boldsymbol{\xi}_P d\boldsymbol{\theta}_{\boldsymbol{P}} \end{aligned} \tag{3-19}$$

La última factorización de la ecuación 3-19 ilustra el acercamiento jerarquico del modelo. Las obervaciones son modeladas independiente condicionalmente en la primera etapa dados los parámetros de los ítems y de las personas, $p(y_{ik} \mid \theta_i, \xi_k)$. Esta es la verosimilitud del modelo que describe la distribución de los datos dados los parámetros de la primera etapa. En una segunda etapa, se especifican las a priori para los parámetros de la primera etapa. Las a priori de la primera etapa describen la heterogeneidad entre individuos $p(\theta_i \mid \boldsymbol{\theta}_P)$ y de la misma forma la heterogeneidad entre ítems $p(\xi_k, \boldsymbol{\xi_P})$. En una tercera etapa, hiperprior son definidas para los parámetros de las densidades prior propuestas.

El problema de integración se expande rápidamente al calcular las medias posteriores marginales para todos los parámetros. De manera que el objetivo es obtener la densidad marginal posterior de los parámetros estructurales del modelo. Esto se realiza a través de integrar la densidad posterior conjunta de todos los parámetros sobre la densidad de los parámetros de perturbación. Una vez la densidad posterior marginal de los parámetros de interés es calculada; inferencias, intervalos posteriores, la moda y media posterior pueden ser calculadas a partir de la densidad marginal posterior. Por ejemplo, el cálculo de la densidad de la habilidad posterior requiere integrar sobre el parámetro del ítem, población y densidad de habilidad de la población. Esto lleva al cálculo de:

$$p(\boldsymbol{\xi}, \boldsymbol{\theta} \mid y) \propto \int\int\int p(y \mid \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\theta} \mid \boldsymbol{\theta}_P) p(\boldsymbol{\theta}_P) p(\boldsymbol{\xi} \mid \boldsymbol{\xi}_P) p(\boldsymbol{\xi}_P) d\boldsymbol{\theta}_P d\boldsymbol{\xi} d\boldsymbol{\xi}_P \tag{3-20}$$

Obtener una aproximación satisfactoria del problema vía integración numérica es una tarea compleja, y los métodos de integración numérica no pueden tratar integrales en dimensiones mayores a 4 [Fox, 2010]. Es por esto que métodos basados en simulación son atractivos por dos razones. Primero, los problemas de integración a menudo involucran distribuciones de probabilidad en la integración, lo que insinúa el uso de métodos de simulación. En segundo lugar, los métodos de simulación son generalmente sencillos de implementar, ya que a menudo se basan en unos pocos principios de simulación y la estructura del problema.

3.2.5. Métodos MCMC

La introducción de los potentes métodos de simulación han hecho posible la aplicación de los métodos bayesianos en varios campos. Los métodos de simulación actuales hacen que las distribuciones posteriores sean accesibles, esto significa que los métodos pueden ser usados para obtener aproximaciones a los momentos posteriores.

Simular valores directamente de la distribución posterior no es posible y debido a los problemas numéricos hasta aquí mencionados, se hace necesario emplear métodos más sofisticados. Una clase de algoritmos conocidos como cadenas de markov montecarlo (MCMC) construyen secuencias que convergen en distribución a la distribución (objetivo) posterior. Posteriormente, promedios muéstrales son calculados para estimar los momentos esperados posteriores[11]. En MCMC, las muestras son generadas a través de una cadena de markov que tiene la distribución objetivo como su distribución estacionaria.

Muestreador de Gibbs

El método más popular de MCMC es el muestreador de Gibbs. Este comienza con una partición de los parámetros o vectores aleatorios de interés en los subvectores $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_Q})$. La densidad posterior conjunta es igual a $p(\boldsymbol{\theta} \mid \mathbf{y})$ y esta es también la densidad objetivo. Un proceso de transición de $\boldsymbol{\theta}^{(m)}$ a $\boldsymbol{\theta}^{(m+1)}$ es definido generando valores en la iteración $m+1$ de la función de densidad de probabilidad condicional de cada subvector,

$$
\begin{aligned}
&\boldsymbol{\theta}_1^{(m+1)} \sim p(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(m)}, \ldots, \boldsymbol{\theta}_q^{(m)}, \mathbf{y}) \\
&\boldsymbol{\theta}_2^{m+1} \sim p(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^{(m+1)}, \boldsymbol{\theta}_3^{(m)}, \ldots, \boldsymbol{\theta}_q^{(m)}, \mathbf{y}) \\
&\vdots \\
&\boldsymbol{\theta}_Q^{m+1} \sim p(\boldsymbol{\theta}_Q \mid \boldsymbol{\theta}_1^{(m+1)}, \ldots, \boldsymbol{\theta}_{Q-1}^{(m+1)}, \mathbf{y})
\end{aligned}
\tag{3-21}
$$

El muestreador de Gibbs genera un proceso de transición, y la forma de las densidades conodicionales y la elección de los bloques caracteriza cada muestreador. Bajo ciertas condiciones de regularidad, la cadena generada de MCMC tiene una densidad estacionaria igual a $P(\theta \mid \mathbf{y})$. Esto significa que $\boldsymbol{\theta}^{(m)}$ converge a $\boldsymbol{\theta}$ en distribución para $m \to \infty$. Para profundizar en las condiciones de regularidad que aseguran un recorrido exitoso para la distribución

[11] Basicamente MCMC es es la integración de montecarlo usando cadenas de Markov

posterior revisar [Fox, 2010, pág 46-47].

Muestreador Metropolis Hastings

Otro método muy empleado dentro de MCMC es el algoritmo Metropolis-Hastings desarrollado por (Metropolis *et al.*, 1953) y generalizado por Hastings 20 años después. El algoritmo M-H generaliza el muestreador de Gibbs ya que ofrece una solución al problema de muestrear desde una distribución condicional, de la cual es difícil muestrear directamente. En un procedimiento de en dos etapas, la superficie de la posterior de interés es explorada a través de una cadena de markov. En el primer paso, un valor candidato es generado de una densidad propuesta que se quiere aproximar a la densidad deseada (objetivo). En el segundo paso, este valor candidato es aceptado o rechazado según un criterio de aceptación.

Técnicamente, en la iteración m un candidato $\boldsymbol{\theta}_q^*$ es generado de una densidad propuesta $q(\boldsymbol{\theta_q} \mid \boldsymbol{\theta}_q^{(m)})$ y la transición de $\boldsymbol{\theta_q^{(m)}}$ a $\boldsymbol{\theta_q^{(m+1)}} = \boldsymbol{\theta_q^*}$ es realizada si

$$u^{(m)} \leq \frac{p(\boldsymbol{\theta}_q^* \mid \mathbf{y})/q(\boldsymbol{\theta}_q^* \mid \boldsymbol{\theta}_q^{(m)})}{p(\boldsymbol{\theta}_q^{(m)} \mid \mathbf{y})/q(\boldsymbol{\theta}_q^{(m)} \mid \boldsymbol{\theta}_q^{(*)})} \tag{3-22}$$

en donde $u^{(m)}$ es la m-ésima observación de una variable aleatorio U que es uniformemente distribuida en el intervalo [0,1]. Si el valor propuesto no es aceptado, la cadena permanece en su valor actual. Dos casos especiales deben ser considerados. El primero, cuando la densidad propuesta es simétrica, $q(\boldsymbol{\theta}_q^* \mid \boldsymbol{\theta}_q^{(m)}) = q(\boldsymbol{\theta}_q^{(m)} \mid \boldsymbol{\theta}_q^*)$, los términos se cancelan en la ecuación 3-22 y se obtiene el algoritmo de (Metropolis *et al.*, 1953). En el segundo caso, cuando la densidad propuesta no depende de valores previos $q(\boldsymbol{\theta}_q \mid \boldsymbol{\theta}_q^{(m)}) = q(\boldsymbol{\theta}_q)$ se obtiene la cadena independiente de Metrópolis.

Monte Carlo Hamiltoniano

En contraste a la caminata aleatoria del algoritmo de Metropolis, en donde el valor propuesto no está relacionado a la distribución objetivo, el método Monte Carlo Hamiltoniano MCH propone un valor calculando la trayectoria acorde a a la dinámica Hamiltoniana que tienen en cuenta la distribución objetivo [Gao, 2018], la especificación de la dinámica Hamiltoniana puede ser consultada en [Steve Brooks, 2012, Cap 5].

Retomando el objetivo inicial de MCMC, se quiere simular una muestra de una densidad objetivo $p(\theta)$. MCH introduce variables de momento auxiliares ϕ y muestrea de una densidad conjunta $p(\theta, \phi)$ que define una función hamiltoniana como en la ecuación 3-23, en donde $U(\theta)$ es la energía potencial y $K(\phi)$ es alguna constante llamada energía cinética.

$$H(\theta,\phi) = -\log\, p(\theta,\phi) = -\log\, p(\theta) - \log\, p(\phi) = U(\theta) + K(\phi) \tag{3-23}$$

Comenzando a partir del valor de los parámetros al final la t-1 ésima iteración, θ^{t-1}, un nuevo valor θ^* es propuesto dos pasos antes de ser evaluado en el paso de aceptación. Primero, un valor ϕ^{t-1} para el momento es generado de la distribución gausiana multivariada N(0,$\mathbf{\Sigma}$). Luego el sistema conjunto $(\theta^{t-1},\phi^{t-1})$ es involucrado a través del siguiente método (ecuación 3-24) para L pasos y así conseguir el valor propuesto (θ^*,ϕ^*):

$$\begin{aligned} \phi_i^{t-1+\varepsilon/2} &= \phi_i^{t-1} + (\varepsilon/2)\frac{\partial \text{log} p(\theta^{t-1})}{\partial \theta_i} \\ \theta_i^{t-1+\varepsilon} &= \phi_i^{t-1} + \varepsilon[\mathbf{\Sigma}^{-1}\phi^{t-1+\varepsilon/2}]_i \\ \phi_i^{t-1+\varepsilon} &= \phi_i^{t-1+\varepsilon/2} + (\varepsilon/2)\frac{\partial \text{log} p(\theta^{t-1+\varepsilon})}{\partial \theta_i} \end{aligned} \tag{3-24}$$

se debe notar que $\theta^* = \theta^{t-1+\varepsilon L}$. Si no hubieran errores numéricos en el procedimiento de la ecuación 3-24, se aceptaría directamente (θ^*,ϕ^*). Sin embargo, siempre hay errores dado el tamaño del paso no cero. Por lo tanto, se debe realizar un paso de aceptación de metrópolis con una tasa de aceptación como sigue:

$$\min\left\{1, \exp[H(\theta^{t-1},\phi^{t-1}) - H(\theta^*,\phi^*)]\right\}$$

En la literatura se sugiere que MCH es eficiente cuando su tasa de aceptación es aproximadamente 65 %.

El muestreador No-U-Turn (NUTS)

En MCH hay tres parámetros de ajuste, la matriz de masa $\mathbf{\Sigma}$, el tamaño del paso ε y el número de pasos $\mathbf{L}$. Si ε es es muy grande, la trayectoria resultante será imprecisa y con alta tasa de rechazo. Si ε es demasiado pequeño o $\mathbf{L}$ es demasiado grande, muchas iteraciones serán llevadas a cabo y el tiempo por iteración será extenso.

NUTS es una extensión de MCH en donde el ajuste de pasos del algoritmo es realizado de forma automática empleando un algoritmo de construcción de arboles. Una sola trayectoria en NUTS es construida acumulando los pasos iterativamente, es decir; el primer paso de la iteración tiene dos pasos, el segundo tiene cuatro, el tercero ocho y de esta forma se duplican los pasos hasta que la trayectoria da vuelta en si misma (U TURN) o diverge [Monnahan et al., 2017], el número de doblaje de pasos se conoce como la profundidad del árbol y el tamaño de cada paso se determina adaptándolo durante la fase de calentamiento con una tasa de aceptación objetivo.

NUTS hace posible presentar inferencia posterior bayesiana de forma potencialmente eficientemente, en un una gran clase de modelos complejos con alta dimensionalidad, con mínima intervención humana. El objetivo de los autores [Hoffman and Gelman, 2014] es que NUTS permitirá a los investigadores preocuparse menos en la fase de ajuste de los modelos y así estos puedan gastar más tiempo en el desarrollo y diagnostico de los modelos. NUTS ha sido implementado en C++ como parte del nuevo paquete de inferencia bayesiana llamado Stan[12].

3.2.6. Estimación de los parámetros en los modelos TRI

En el proceso de estimación de los parámetros de los modelos TRI, algunos de los parámetros son de interés (estructurales) y otros no (parámetros de perturbación). El objetivo principal aquí es obtener la densidad posterior marginal de los parámetros estructurales integrando la densidad conjunta posterior de todos los parámetros sobre la densidad de los parámetros de perturbación.

Un método alternativo que evita los problemas de integración fue propuesto por Swaminathan y Gifford (1982). Un procedimiento de Newton Raphson es usado para obtener las estimaciones que maximizan la posterior conjunta. Tanto los parámetros estructurales como de perturbación que maximizan la distribución conjunta posterior son considerados estimaciones bayesianas. Sin embargo, estas estimaciones posteriores conjuntas se consideran inferiores a las estimaciones posteriores marginales.

En la propuesta de de Bock y Atkin (1981) se emplea el algoritmo de esperanza máxima EM, en donde el interés se centra en los parámetros de los ítems y los parámetros de individuos son tratados como parámetros de perturbación como se observa en la ecuación 3-23. Esta expresión maximizada resulta en la estimación de los parámetros de los ítems a través de máxima verosimilitud marginal.

$$\max\, l(\mathbf{y};\xi) = \int l(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3-25}$$

En el siguiente paso se obtienen estimadores por medio de estadísticos suficientes, a partir de los cuales se generan pseudo datos[13], los cuales son actualizados cada vez en el paso E dadas las estimaciones de los parámetros de los ítems. En el paso M, las estimaciones de los parámetros de los ítems son obtenidas dados los pseudo datos actualizados. Ambos pasos se

[12]En el Anexo A se encuentra una introducción al software STAN

[13]Estos corresponden al número de individuos que se esperan tengan cierto nivel de la variable latente, y de estos el número que responda correctamente

repiten hasta que un nivel de convergencia es alcanzado.

En un enfoque jerarquico bayesiano, las inferencias para los parámetros de los ítems son realizadas a partir de la posterior marginal como se ve en la expresión 3-24,

$$p(\boldsymbol{\xi} \mid \mathbf{y}) \propto \int \int \int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi}) p(\boldsymbol{\theta} \mid \boldsymbol{\theta}_P) p(\boldsymbol{\xi} \mid \boldsymbol{\xi}_P) p(\boldsymbol{\xi}_P) p(\boldsymbol{\theta}_P) d\boldsymbol{\theta} d\boldsymbol{\theta}_P d\boldsymbol{\xi}_P \tag{3-26}$$

Sin embargo, para realizar una estimación desde el enfoque jerárquico bayesiano, integrales de altas dimensiones necesitan ser evaluadas. Por lo tanto, métodos computacionales son requeridos para soportar el cálculo de densidades marginales posteriores y sus respectivos análisis.

4. Metodología

4.1. El instrumento y los datos

Con el objetivo de construir un modelo que permita evaluar el perfil de riesgo de las personas naturales en el mercado financiero colombiano, se diseñó un instrumento (Anexo B): compuesto de 18 ítems, cada uno con tres respuestas posibles ordenadas (1, 2 y 3), en donde la categoría uno indica un nivel bajo de riesgo de inversión y la categoría tres un nivel alto. El instrumento fue desarrollado, recopilando e investigando las tipologías de preguntas que realizan las principales instituciones financieras del país (sección 2.3) en el proceso de evaluación e inscripción de sus clientes inversionistas. Esta revisión permitió identificar tres tipologías principales de preguntas (ítems): etapa económica del individuo al punto de evaluación (ítems B, I, K, J), comportamiento de elección en diferentes escenarios de incertidumbre financiera (ítems A, C, D, E, F, G, H) y conocimiento y experiencia en productos financieros (ítems L, M , N , O , P , Q ,R). Esta etapa fue exploratoria y de propuesta inicial de ítems evaluados, ya que en la sección 5.1 se explica la metodología de selección final de ítems.

El objetivo inicial fue aplicar el instrumento de forma presencial, pero debido a las restricciones surgidas con el brote de Covid-19 a finales de febrero en Colombia, prudentemente la aplicación se realizó de forma electrónica. Considerando que no hay un dictamen definitivo en la literatura para el tamaño de muestra en modelos TRI, en varios trabajos se pueden encontrar lineamientos generalmente aceptados. Primero, el tamaño de muestra debe aumentar con la complejidad del modelo[1] [Edelen and Reeve, 2007] y segundo, tamaños de muestra de hasta cien observaciones a menudo son adecuados para estimar parámetros estables del modelo de Rasch, ver trabajo de [Linacre J, 1994]. Es por esto que en la propuesta de este trabajo se fijo el objetivo de 100 observaciones, pero debido a la difusión que tuvo el instrumento se llegaron a obtener 144 respuestas, de las cuales el 69 % corresponden a estudiantes de maestría en estadística, economía o finanzas, 10 % a estudiantes de especialización en estadística o finanzas y 21 % a estudiantes de pregrado en estadística o economía.

[1] La complejidad del modelo aquí se refiere al número de parámetros.

4.2. El modelo propuesto

Teniendo en cuenta que uno de los objetivos principales en este trabajo es obtener una medida unidimensional[2] que permita modelar el perfil de riesgo de inversión, los modelos TRI por construcción son caracterizados por unidimensionalidad [Caviezel et al., 2011] e independencia local como se explica en la sección 3.1, ecuación 3-2. Además, teniendo en cuenta la naturaleza polítomica[3] de los datos obtenidos por medio del instrumento , el trabajo de (Muraki, 1992) propone y demuestra que el modelo Generalized Partial Credit Model es bastante flexible para este tipo de datos y es adecuado cuando los ítems no poseen el mismo nivel de discriminación, esta versión del modelo como se observa en la ecuación 4-1 posee la misma estructura que el Partial Credit Model pero se añade un parámetro de discriminación α_k por ítem.

$$P(Y_{ik} = c \mid \theta_i, \boldsymbol{\kappa}_k, \alpha_k) = \frac{exp \sum_{l=1}^{c}(\alpha_k \theta_i - \kappa_{k,l})}{\sum_{r=1}^{C_k}(exp \sum_{l=1}^{r}(\alpha_k \theta_i - \kappa_{k,l}))}, \tag{4-1}$$

En donde $P(Y_{ik} = c \mid \theta_i, \boldsymbol{\kappa}_k, \alpha_k)$ es la probabilidad que el individuo i con nivel de habilidad θ_i de una respuesta en una categoría particular $c(c = 1, ..., C_k)$ para el ítem k, con $\kappa_{k,l}$ como el parámetro de dificultad de paso del ítem. En contraste al modelo TRI de respuesta binaria, acá la interpretación del parámetro de discriminación α_k indica el nivel de cambio de las respuestas categóricas (1,2 y 3) al interior del ítem en relación a un cambio en el rasgo latente θ.

El modelo Bayesiano jerárquico es completado especificando distribuciones a priori para todos lo parámetros e hiperparámetros. A continuación se expone y se argumenta la elección de dichas distribuciones prior:

$$\begin{aligned}
\theta_i &\sim N(0,1),\\
\alpha_k &\sim N(0,2),\\
\kappa_{k,l} &\sim N(\mu_{\kappa_{k,l}}, \sigma_{\kappa_{k,l}}),\\
\mu_{\kappa_{k,l}} &\sim N(0,1),\\
\sigma_{\kappa_{k,l}} &\sim Cauchy(0,2).
\end{aligned}$$

Ya que se puede multiplicar cada α_k y dividir θ_i,$\kappa_{k,l}$ por una misma constante sin cambiar el valor de la función de verosimilitud, se deben fijar los parámetros de escala y habilidad para que θ_i sea identificable, es por esto que se asume $\theta_i \sim N(0,1)$ como se aconseja en el trabajo de [Rangel, 2019]. Para elegir la distribución a priori de α_k es necesario recordar el supuesto de monotonicidad [4], por lo cual α_k debe ser restringido a ser positivo [Fox, 2010], es por esto

[2]Unidimensionalidad significa que un solo constructo es medido (por ejemplo el rasgo latente representado por un escala)

[3]ítems en donde la categoría de respuestas pueden ser ordenadas de acuerdo a nivel habilidad

[4]Este supuesto implica que $P(Y_{ik} \mid \theta_i)$ es no decreciente en θ_i

que acá se asume $\alpha_k \sim N(0,2)$ truncado en cero. En el caso del parámetro $\kappa_{k,l}$, teniendo en cuenta que por definición este se encuentra en la misma escala que θ_i, se asume normal con el hiper parámetro $\mu_{\kappa_{k,l}} \sim N(0,1)$ y $\sigma_{\kappa_{k,l}} \sim Cauchy(0,2)$ ya que el uso frecuente de la familia gamma inversa para parámetros de varianza ha mostrado problemas, el trabajo de (Gelman (2006). Citado en Molina, 2016, p.6) demuestra que el uso de la distribución Half Cauchy como a priori para los parámetros de varianza se comporta bien ya que asintoticamente es no informativa y se desempeña bien alrededor de cero [Monila, 2016].

5. Estimación y Resultados

Antes de llevar a cabo el proceso de estimación, un requerimiento fundamental previo en la ejecución de modelos TRI es la verificación de unidimensionalidad[1] de los ítems en relación al rasgo latente de interés en la investigación [Brentari and Golia, 2007]. Para evaluar esta propiedad, se decidió evaluar la dirección de los ítems propuestos[2] a partir de los coeficientes de los vectores propios en un análisis de componentes principales (ACP). En la figura **5-1** se presenta el resultado de los coeficientes para los dos primeros componentes que explican el 33 % de la varianza.

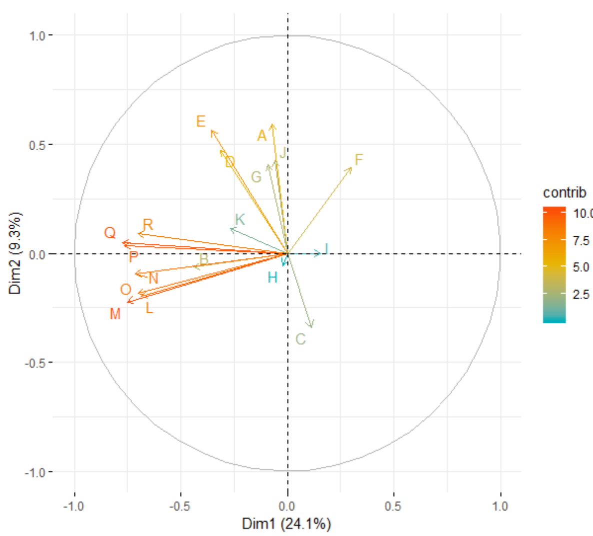

Figura 5-1.: Análisis de componentes principales para el conjunto de ítems inicial.

En la figura **5-1** se observa que los coeficientes de los vectores propios para los ítmes I, C y F tienen una dirección positiva en el primer componente en contraste al resto de ítems, lo cual sugiere revisar el orden planteado de las categorías de respuestas para estos y evaluar si tiene sentido invertir dicho orden. La evaluación del orden reveló que para el ítem C, el orden planteado se encontraba invertido, para los ítems F e I se diagnosticó ambigüedad en el planteamiento de la pregunta, por lo cual estos dos (F,I) fueron descartados en el proceso de estimación. Además, el ítem H tampoco se considera, debido su baja contribución en la

[1]El concepto de unidimensionalidad se define con frecuencia como un rasgo latente único que puede ser explicado a partir del rendimiento de los ítems que forman un cuestionario[Brentari and Golia, 2007]

[2]Los ítems iniciales de partida fueron (A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R)

matriz de pesos con un valor de $-0{,}9\,\%$. Finalmente, a parir del análisis ACP, se considera pertinente no tener en cuenta los ítem I, F y H en el proceso de estimación, y para el ítem C, incluirlo en la estimación pero con el orden invertido en las categorías de respuestas.

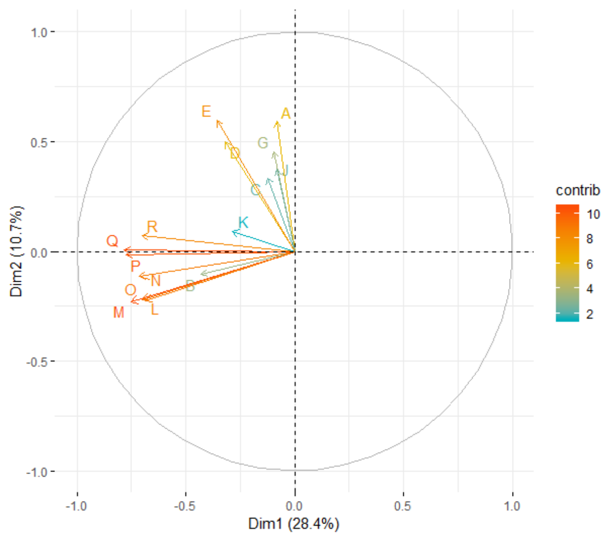

Figura 5-2.: Análisis de componentes principales para el conjunto de ítems inicial menos I,F,C y orden invertido para C.

El gráfico **5-2** refleja que los 15 ítems finales (A,B,C,D,E,G,J,K,L,M,N,O,P,Q y R) tienen un sentido común en términos de dirección, lo cual apoya la verificación de undimensionalidad del constructo y permite validar el sentido de las categorías de los ítems en relación al rasgo lantente evaluado que es perfil de riesgo de inversión de los individuos.

5.1. Algoritmo de estimación

El software usado en este trabajo (Stan) para el proceso de estimación emplea el muestreador no-U-turn sampler explicado en el marco teórico de este trabajo (sección 3.2.5), el cual es una generalización dentro de la familia de algoritmos de metrópolis, presentando múltiples pasos por iteración para moverse más eficientemente a través de la distribución posterior. En [Hoffman and Gelman, 2014] se encuentra el desarrollo formal y aplicación del algoritmo en el lenguaje C++ para el nuevo paquete de inferencia Bayesiana Stan. Para usar Stan, el usuario debe escribir un *programa stan* que contiene un modelo bayesiano en un lenguaje conveniente cuyo código luce como notación estadística, y que permite calcular directamente la log-densidad posterior.

5.2. Estimación de parámetros del modelo GPCM

La función *stan* del paquete **Rstan** (Anexo A) se empleo en el software R para obtener la información y estadísticos de sumarización de la distribución posterior de todos los parámetros del modelo. La media posterior y la desviación estándar pueden ser usadas como estimaciones puntuales de los parámetros de discriminación y sus errores estándar.

En la tabla 5-1 presenta los resultados para los parámetros de discriminación. El ítem que presenta mejor capacidad de discriminación es el 14 (Q) con un valor de 2.7, correspondiente a la evaluación personal de experiencia en inversiones, seguido del ítem 13 (P) con un valor de 2.55 relacionado con el conocimiento de los productos financieros más sofisticados como los son derivados o activos de renta variable, estos resultados indican que dada su capacidad de discriminación, ítems relacionados con el conocimiento o experiencia financiera, deben ser imprescindibles en los cuestionarios empleados por las entidades para evaluar el perfil de riesgo de sus clientes, además de tener mucho sentido tener en cuenta la evaluación personal de conocimientos al momento de realizar una inversión.

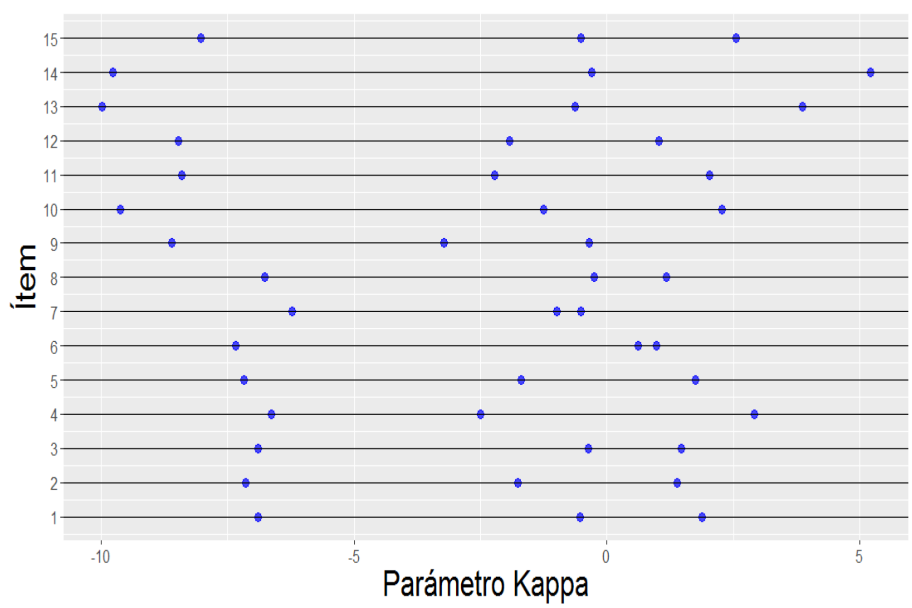

Figura 5-3.: Parámetros de dificultad estimados

En contraste, el ítem con menor capacidad de discriminación con un valor de 0.15 fue el 6 (G) correspondiente al comportamiento de elección bajo un escenario con diferentes niveles de relación riesgo-ganancia. Esta pregunta (G) es interesante ya que pone a prueba en el encuestado conocimientos básicos de probabilidad e incluso valor esperado, lo cual puede indicar que cuando los individuos no tienen bases matemáticas suficientes, sus decisiones pueden no ser objetivas y por lo tanto en situaciones similares a la del ítem (G), tienden

Tabla 5-1.: Estadísticas descriptivas de los parámetros de discriminación

$\boldsymbol{\alpha_i}$	$\boldsymbol{\mu}$	**SE**	$\boldsymbol{\sigma}$	**25 %**	**50 %**	**75 %**	**97.50 %**	**neff**	**Rhat**
α_1	0.16	0.00	0.11	0.07	0.14	0.23	0.41	1847	1.00
α_2	0.74	0.00	0.20	0.60	0.73	0.87	1.16	7372	1.00
α_3	0.19	0.00	0.12	0.10	0.18	0.27	0.47	2027	1.00
α_4	0.72	0.00	0.26	0.54	0.71	0.88	1.27	5317	1.00
α_5	0.60	0.00	0.20	0.46	0.59	0.73	1.00	6960	1.00
α_6	0.15	0.00	0.10	0.07	0.13	0.21	0.38	2032	1.00
α_7	0.16	0.00	0.10	0.08	0.14	0.22	0.39	2834	1.00
α_8	0.35	0.00	0.14	0.26	0.35	0.44	0.65	4109	1.00
α_9	2.03	0.01	0.41	1.74	2.00	2.29	2.93	2723	1.00
α_{10}	2.32	0.01	0.44	2.01	2.28	2.59	3.27	2669	1.00
α_{11}	1.93	0.01	0.36	1.68	1.91	2.16	2.72	4251	1.00
α_{12}	1.80	0.01	0.35	1.55	1.78	2.02	2.54	3425	1.00
α_{13}	2.55	0.01	0.48	2.21	2.51	2.85	3.57	2758	1.00
α_{14}	2.70	0.01	0.51	2.35	2.67	3.02	3.80	2008	1.00
α_{15}	1.74	0.01	0.34	1.50	1.72	1.95	2.47	4062	1.00

Tabla 5-2.: Estadísticas descriptivas de los parámetros de dificultad

Parámetros	$\boldsymbol{\kappa_{k,l}}$	Parámetros	$\boldsymbol{\kappa_{k,l}}$	Parámetros	$\boldsymbol{\kappa_{k,l}}$
$\kappa_{1,1}$	-6.90	$\kappa_{6,1}$	-7.34	$\kappa_{11,1}$	-8.41
$\kappa_{1,2}$	-0.53	$\kappa_{6,2}$	0.62	$\kappa_{11,2}$	-2.23
$\kappa_{1,3}$	1.88	$\kappa_{6,3}$	0.98	$\kappa_{11,3}$	2.02
$\kappa_{2,1}$	-7.15	$\kappa_{7,1}$	-6.23	$\kappa_{12,1}$	-8.48
$\kappa_{2,2}$	-1.76	$\kappa_{7,2}$	-1.00	$\kappa_{12,2}$	-1.92
$\kappa_{2,3}$	1.38	$\kappa_{7,3}$	-0.52	$\kappa_{12,3}$	1.02
$\kappa_{3,1}$	-6.90	$\kappa_{8,1}$	-6.77	$\kappa_{13,1}$	-9.98
$\kappa_{3,2}$	-0.37	$\kappa_{8,2}$	-0.25	$\kappa_{13,2}$	-0.63
$\kappa_{3,3}$	1.47	$\kappa_{8,3}$	1.17	$\kappa_{13,3}$	3.87
$\kappa_{4,1}$	-6.64	$\kappa_{9,1}$	-8.61	$\kappa_{14,1}$	-9.78
$\kappa_{4,2}$	-2.50	$\kappa_{9,2}$	-3.22	$\kappa_{14,2}$	-0.31
$\kappa_{4,3}$	2.91	$\kappa_{9,3}$	-0.35	$\kappa_{14,3}$	5.21
$\kappa_{5,1}$	-7.18	$\kappa_{10,1}$	-9.62	$\kappa_{15,1}$	-8.03
$\kappa_{5,2}$	-1.70	$\kappa_{10,2}$	-1.25	$\kappa_{15,2}$	-0.51
$\kappa_{5,3}$	1.75	$\kappa_{10,3}$	2.27	$\kappa_{15,3}$	2.55

a elegir posiciones diferentes a su verdadero apetito de riesgo y es por esto que el ítem no presenta una buena capacidad de discriminación.

Respecto a los parámetros de dificultad $\kappa_{k,l}$, se estimaron 45 de estos, correspondientes a tres niveles por cada uno de los 15 ítems. En la tabla 5-2 y el el gráfico **5-3** se presenta el resumen de las estadísticas a posteriori, allí se observa el correcto ordenamiento de los parámetros de dificultad, es decir que la primera categoría $\kappa_{k,1}$ en cada ítem posee el menor valor de dificultad y la última categoría $\kappa_{k,3}$ posee el mayor valor[3]. Este ordenamiento indica que para que un individuo pueda pasar de la segunda categoría de respuesta a la tercera categoría, se requiere un nivel de habilidad (trazo latente o perfil de riesgo) más alto en contraste al nivel habilidad requerido para saltar de la primera categoría a la segunda al interior de todos los ítem, esta característica se puede presentar de forma gráfica como se ve en la figura **5-4** para el caso del ítem 14 (G). Según el autor de este modelo [Masters, 1982], estas curvas se denominan curvas características categóricas del ítem (CCCI) y representan la probabilidad de respuesta en cada categoría a través el dominio del rasgo latente θ.

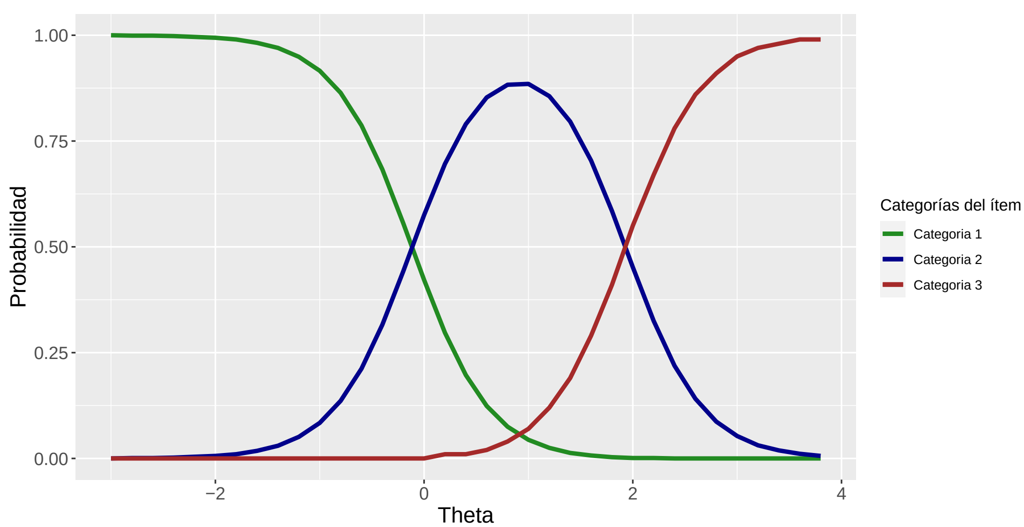

Figura 5-4.: CCCI ítem Q (14)

La figura **5-4** ilustra muy bien el comportamiento de la probabilidad de elección de las categorías al interior del ítem. La curva de color verde representa la probabilidad de elección de la primera categoría del ítem (experiencia limitada en inversiones), la cual tiende a disminuir a medida que aumenta el rasgo latente (perfil de riesgo). Se observa que a medida que disminuye la probabilidad de la categoría uno y aumenta el valor de θ, empieza a aumentar la

[3]Este resultado tiene sentido con el diseño de la encuesta para cada ítem, es decir que los individuos que responden en la tercera categoría , se esperan tengan un perfil de riesgo de inversión más alto en contraste a individuos que responden la primera o segunda categoría.

probabilidad para la segunda categoría (Experiencia moderada en inversiones) y de la misma forma cuando esta empieza a bajar, la probabilidad de la tercera categoría (Experiencia suficiente en inversiones) empieza a aumentar, hasta alcanzar el límite del valor uno.

El ítem que presenta el nivel de dificultad más alto en la categoría 3 es el 14 (Q) con un valor de 5.21, lo cual concuerda con el resultado de los parámetros de discriminación al ser este el que más discrimina, este resultado tiene sentido con la estimación de parámetros de dificultad ya que alcanzar la categoría 3 en este ítem supone un perfil de riesgo más elevado (rasgo latente o habilidad medida en este estudio).

Tabla 5-3.: Estadística descriptiva del trazo latente

Mínimo	Cuartil 1	Mediana	Promedio	Cuartil 3	Máximo
-1.940	-0.596	-0.024	0.012	0.580	2.584

La tabla 5-3 presenta la estadística descriptiva de los promedios de habilidad para los 144 individuos evaluados y el gráfico **5-5** presenta los intervalos de credibilidad del 95 %. Se evidencia que el promedio se encuentra alrededor de cero acorde a la distribución a priori seleccionada. El individuo con trazo menor posee un valor de -1.9 y el máximo posee un valor de 2.58. A continuación, se presenta un gráfico que compara los resultados desde la teoría clásica de test[4] y los promedios obtenidos de la estimación para el trazo latente bajo el modelo propuesto. La figura **5-6** permite observar la dispersión del puntaje clásico de *test* y los promedios de los parámetros de habilidad θ_i, se evidencia una relación lineal con un coeficiente de pearson de 0.93. Este gráfico permite ver una de las bondades de emplear un modelo TRI para la estimación de los trazos latentes, ya que el dominio del parámetro θ se mantiene en un rango, sin importar el número de ítems evaluados.

Ya que uno de los objetivos principales de este trabajo es establecer una posible clasificación de los individuos desde el marco de TRI en niveles de riesgo de inversión, una propuesta de clasificación para los individuos se puede realizar tomando el rango intercuartílico del trazo latente, ya que permite tener tres niveles de riesgo como normalmente se requiere en el sector financiero (aversos al riesgo, riesgo moderado y amantes al riesgo). De esta forma, se propone la siguiente clasificación para los individuos como se observa en la tabla 5-4 y se ve gráficamente en la figura **5-5**.

[4]En la teoría clásica de test (TCT) los puntajes asignados para cada individuo como medida del razgo latente, dependen del número de pregunta presentes en el cuestionario, es por esto que el dominio de los puntajes clásicos en este ejemplo van desde cero hasta 40, y si el cuestionario empleará más ítem, el dominio sería mayor a 40

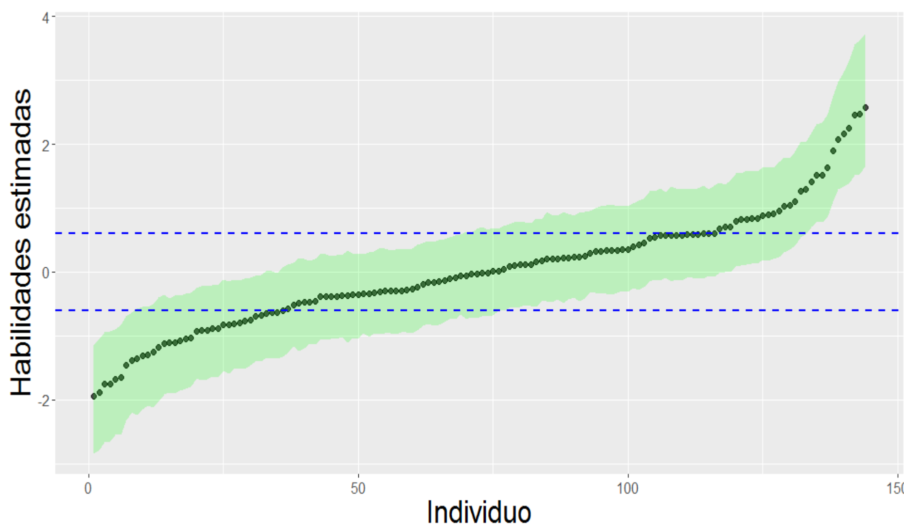

Figura 5-5.: Bandas de credibilidad del 95 % para el parámetro θ_i

Tabla 5-4.: Clasificación de individuos según rango intercuartílico de las medias posteriores para el parámetro de habilidad.

Parámetro	Regla	Rango	Nivel
	<Q1	$(-\infty$, -0.6)	**Averso al riesgo**
	>= Q1 y <= Q3	[-0.6, 0.6]	**Riesgo Moderado**
θ	>Q3	(0.6, $+\infty$)	**Amante al riesgo**

5.3. Evaluación de ajuste del modelo

5.3.1. Diagnostico de convergencia de las cadenas de Markov

Un diagnostico de convergencia es un método para evaluar hasta qué punto se debe ejecutar una cadena de markov para llegar a obtener observaciones de la distribución estacionaria. Normalmente se evalúa graficamente la evolución de las cadenas de markov generadas en el software Stan y el criterio de convergencia Gelman-Rubin $\hat{R}$ (R hat)[5].

El gráfico **5-7** presenta el comportamiento de las cuatro cadenas generadas para la estimación de los parámetros de habilidad θ_6,θ_{117},θ_{250}, en donde se puede inferir una buena convergencia ya que los valores presentan un comportamiento estacionario alrededor del cero.

[5]El R hat es un estadístico que se emplea para el diagnostico de convergencia, el cual compara la varianza de las estimaciones de los parámetros del modelo entre cadenas y dentro de las cadenas, si las cadenas no se han estabilizado bien, los resultados del $\hat{R}$ serán mayores a uno, una recomendación es usar cuatro cadenas y emplear las muestras para el análisis cuando el $\hat{R}$ sea menor a 1.05

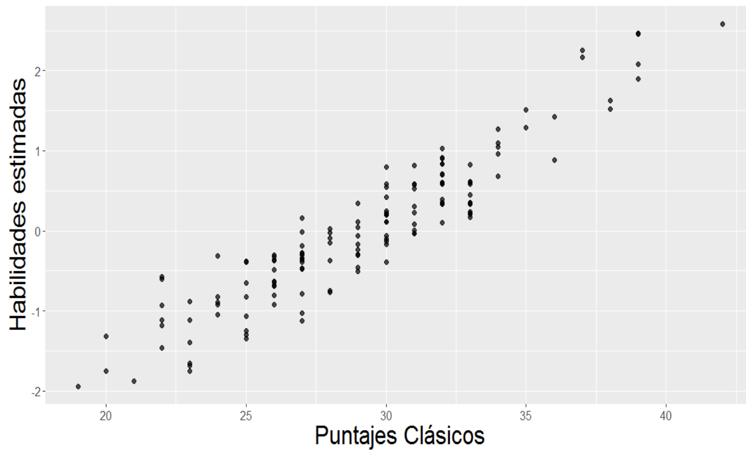

Figura 5-6.: Correlación puntaje Clásico y θ_i estimados

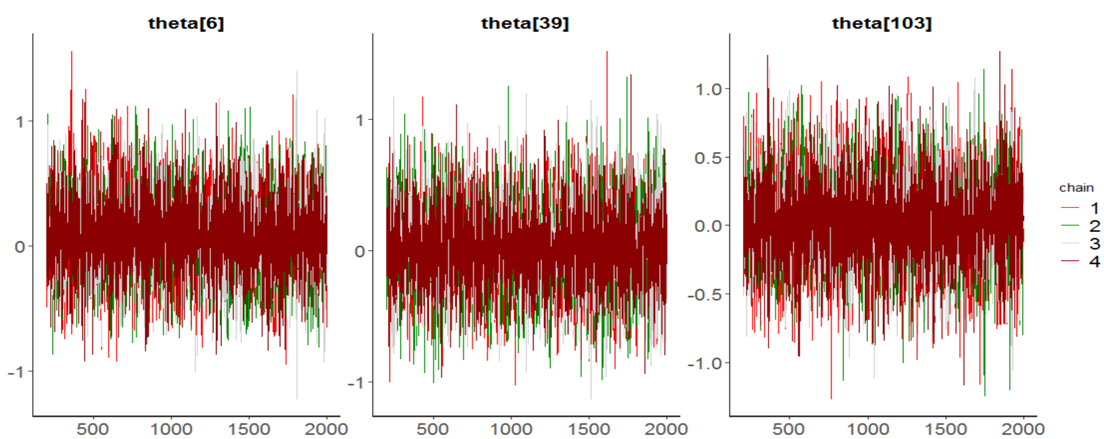

Figura 5-7.: Muestras de las cadenas estimando el trazo latente

Tabla 5-5.: Estadística descriptiva de Rhat del trazo latente

Mínimo	Cuartil 1	Mediana	Promedio	Cuartil 3	Máximo
0.9995	0.9997	0.9999	1.0000	1.0001	1.0012

La tabla 5-5 presentan las estadísticas descriptivas del Rhat para los 144 parámetros de habilidad de los individuos. Se puede concluir que se presenta convergencia ya que la media es uno y la mediana está muy cerca a uno.

La figura **5-8** permite observar el dominio de todos los valores $\hat{R}$ correspondientes a todos los parámetros calculados en el modelo. La columna *theta* θ se encuentra más poblada ya que se tienen los valores $\hat{R}$ para 144 parámetros de habilidad, para el parámetro *kappa* κ de dificultad se tienen 45 valores y para el parámetro de discriminación se tienen 15 valores

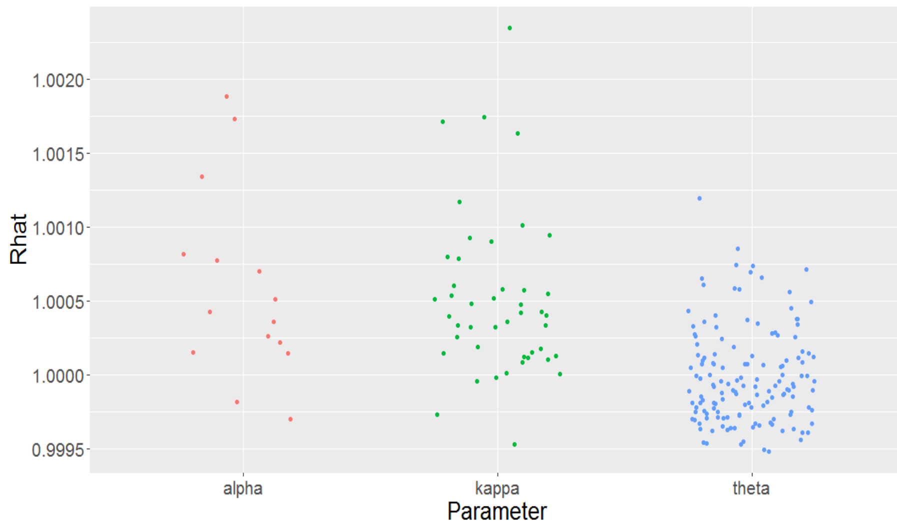

Figura 5-8.: Muestras de las cadenas estimando el trazo latente

correspondientes a 15 ítems. A partir de este gráfico se puede concluir convergencia de las cadenas de markov para todos los parámetros del modelo ya que todos tienen un valor menor a 1.1.

5.3.2. Ajuste del modelo

Una metodología poderosa para la evaluación de modelos Bayesianos se basa en la evaluación predictiva. La idea es generar datos predictivos a partir del modelo estimado, y luego comparar estos con los datos observados [Fox, 2010, Pág 107]. Los estadísticos empleados para llevar acabo la comparación son llamados en la literatura medidas de discrepancia o test estadísticos.

Si se considera un modelo TRI M con parámetros $(\boldsymbol{\theta}, \boldsymbol{\xi})$. Bajo la hipótesis nula, se asume que los datos respuesta están condicionalmente distribuidos como $p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi})$ y los parámetros desconocidos tienen una distribución a priori $p(\boldsymbol{\theta} \mid \boldsymbol{\xi})$, sin tener una hipótesis alternativa especifica, el interés se centra en un estadístico, denotado por $T(\boldsymbol{y})$, para verificar la compatibilidad del modelo con los datos observados $\boldsymbol{y}_{obs}$. De forma que un p-valor bayesiano puede ser calculado como:

$$p_0(\boldsymbol{y}_{obs}) = P(T(\boldsymbol{Y}) \geq T(\boldsymbol{y}_{obs}) \mid M) \tag{5-1}$$

Con el fin de evaluar el ajuste del modelo PCM propuesto, se calculó el P-valor Bayesiano, construyendo el siguiente algoritmo en el software Stan:

1. Se obtiene una muestra $\boldsymbol{\theta}^k$ de la posterior $p(\boldsymbol{\theta} \mid \mathbf{z})$.

2. Se obtiene una réplica $\mathbf{x}^{(k)}$ a partir de la verosimilitud $f(\mathbf{x} \mid \boldsymbol{\theta}^{(k)})$ de la misma forma como se obtuvo la muestra de $\mathbf{x}$.
3. Se calculan las discrepancias $T(\mathbf{x}^{rep} \mid \boldsymbol{\theta}^{(k)})$ y $T(\mathbf{x} \mid \boldsymbol{\theta}^{(k)})$.
4. Se calcula la estadística $1_{\{T(\mathbf{x}^{rep}|\boldsymbol{\theta}^{(k)})>T(\mathbf{x}|\boldsymbol{\theta}^{(k)})\}}$
5. Finalmente el p valor bayesiano es: $p = m^{-1} \sum_k 1_{\{T(\mathbf{x}^{rep}|\boldsymbol{\theta}^{(k)})>T(\mathbf{x}|\boldsymbol{\theta}^{(k)})\}}$, en donde m es el número de muestras de la estadística $1_{\{T(\mathbf{x}^{rep}|\boldsymbol{\theta}^{(k)})>T(\mathbf{x}|\boldsymbol{\theta}^{(k)})\}}$.

Después de construir el algoritmo en el software Stan (Anexo A), se obtuvieron dos criterios de ajuste respecto a los datos observados, la primera medida fue el p valor bayesiano para el ajuste global del modelo y la segunda medida fue un p - valor por cada ítem evaluado.

Tabla 5-6.: Resultados de ajuste (p valor bayesiano) para los ítems

Criterio	P valor Bayesiano	Criterio	P valor Bayesiano	Criterio	P valor Bayesiano
Item 1	0.41	Item 6	0.41	Item 11	0.49
Item 2	0.43	Item 7	0.38	Item 12	0.47
Item 3	0.38	Item 8	0.39	Item 13	0.56
Item 4	0.46	Item 9	0.49	Item 14	0.61
Item 5	0.44	Item 10	0.52	Item 15	0.49

En la tabla 5-6 se observan los resultados del p-valor para los 15 ítem considerados en el estudio, todos los p-valores se encuentran alrededor del 0.5 y no superan 0.9 o son menores a 0.1 [Fox, 2010], lo cual permite concluir que los 15 ítems pasan la prueba de bondad de ajuste. Finalmente, considerando que como ya se mencionó bajo la hipótesis nula, se asume que los datos respuesta están condicionalmente distribuidos como $p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi})$, el valor para el ajuste global del modelo fue de 0.33, lo que permite no rechazar la hipotsís nula y concluir que el modelo propuesto en este trabajo GPCM con su denominación original en inglés *Generalized Partial Credit Model* se ajusta bien a los datos.

6. Conclusiones y recomendaciones

6.1. Conclusiones

Este trabajo identificó falencias teóricas en el desarrollo y aplicación de los instrumentos que emplean las principales instituciones financieras del país para crear una medida relacionada con el perfil de riesgo de los inversionistas. Esta medida según la normatividad, es un insumo necesario para la clasificación de los individuos en niveles de riesgo, lo que permite adecuar mejor las estrategias de inversión acorde al perfil de riesgo obtenido. A partir de este hallazgo, surge la propuesta central de este trabajo.

La principal falencia encontrada en los instrumentos usados en el sector financiero, recae en la construcción de medidas en diferentes escalas dependiendo el número de ítem evaluados por las compañías. Este error se potencia ya que además de tener diferente cantidad de ítems, normalmente se asigna una ponderación distinta a cada ítem y posteriormente se suman los valores de las respuestas a cada ítem para obtener un valor final que corresponde al perfil de riesgo. Después de obtener el puntaje final, una segunda falencia fue identificada: cada compañia tiene un criterio subjetivo de clasificación de los individuos sin un rigor teórico, esto acorde a puntos de corte establecidos según el criterio interno de las compañías.

El modelo GPCM (Generalized Partial Credit Model) propuesto en este trabajo, permite abordar las inconsistencias encontradas en la investigación desde el enfoque de los modelos TRI, este marco permite obtener una medida unidimensional que representa la característica de investigación latente, que es el perfil de riesgo o como se menciona en la literatura, el nivel de tolerancia al riesgo de los individuos. Otra bondad de este tipo de modelos es que son caracterizados por obtener medidas invariantes para los parámetros de los ítem e individuos, permitiendo al investigador evaluar la contribución individual de los ítem incluyéndolos u omitiéndolos del cuestionario.

Dentro de los modelos TRI para respuestas politómicas, específicamente el modelo GPCM en constraste al PCM[1] permite establecer discriminación no uniforme a través de los ítems, lo que significa que la aplicación del GPCM permite una mayor comprensión de las características de los ítems y su contribución dentro del cuestionario. Para la aplicación realizada en

[1]El Partial Credit Model PCM posee la misma estructura que el GPCM pero

este trabajo, el ítem que presento mayor discriminación fue el Q (Anexo B) que corresponde a la evaluación personal de la experiencia en inversiones financieras, lo que dio sentido con la estimación de parámetros de dificultad para el ítem, ya que este ítem también presentó el mayor valor de dificultad para alcanzar la categoría tres de respuesta con un valor $\kappa_{14,3}$ de 5.21. Desde el punto de vista financiero, este resultado es adecuado ya que la evaluación personal a la hora de invertir es crucial en la toma de decisiones financieras de largo plazo.

El ítem que presentó menor capacidad de discriminación fue el (G) correspondiente al comportamiento de elección bajo un escenario con diferentes niveles de relación riesgo-ganancia. Lo cual puede indicar que cuando los individuos no tienen bases matemáticas suficientes, sus decisiones pueden no ser objetivas y por lo tanto en situaciones similares a la del ítem (G), tienden a elegir posiciones diferentes a su verdadero apetito de riesgo y es por esto que el ítem no presenta una buena capacidad de discriminación.

Para la estimación de parámetros del modelo, se empleó un algoritmo extensión del MCH (Monte Carlo Hamiltoniano), este es el algoritmo NUTS (No-U-Turn Sampler) implementado en el relativamente nuevo lenguaje Stan para estadística beyesiana por medio del software estadístico R. Este algoritmo es considerablemente más rápido que el muestreador de Gibss y Metropolis, ya que este explora el espacio de parámetros posterior de una forma más eficiente.

El modelo superó la fase de diagnostico en términos de ajuste, por medio de la evaluación predictiva empleando el criterio del p valor bayesiano. El p valor de los ítems determinaron buen ajuste al encontrarse todos alrededor de 0.5, de la misma forma que el p-valor del modelo a nivel global. En términos de convergencias de las cadenas de markov, también se cumplieron con los límites sugeridos en la literatura para garantizar convergencia de las cadenas.

A. Anexo: Introducción al software estadístico bayesiano STAN

Stan nombrado en honor a Stanislaw Ulam uno de los creadores del método de montecarlo es un programa libre de código abierto escrito en el lenguaje de programación C++ (disponible en http://mc-stan.org/) que permite realizar inferencia bayesiana y optimización de modelos especificados por el usuario de forma similar a la escritura estadística en Bugs y Jags. El programa puede ser ejecutado desde la linea de comandos de programas como R Studio, Python, Matlab o Julia, y tiene grandes cualidades en el ajuste de modelos complejos en varias áreas de aplicación.

Un programa Stan para un modelo estadístico está organizado normalmente en una secuencia de bloques, los cuales son: funciones, datos, transformaciones de los datos, parámetros, parámetros transformados, modelo y generación de cuantiles. La presencia de los bloques es opcional dependiendo los requerimientos del investigador. A continuación se presentan los bloques empleados en este trabajo con una explicación introductoria de su función.

- **Bloque de programa: Funciones**

  Las funciones definidas por el usuario deben encontrarse en primer orden en la declaración de bloques, esto debido a que como en el software R las funciones debe ser declaradas antes de ejecutarse. En este caso se declara la función *pcm* que genera un valor real tomando los argumentos y (valor entero), *theta* (valor real) y *kappa* (vector).

```
functions {
real pcm(int y, real theta, vector kappa) {
vector[rows(kappa) + 1] unsummed;
vector[rows(kappa) + 1] probs;
unsummed = append_row(rep_vector(0.0, 1), theta - kappa);
probs = softmax(cumulative_sum(unsummed));
return categorical_lpmf(y + 1 | probs);
}
```

  Esta función en un primer paso crea dos vectores llamados *unsummed* y *probs* del tamaño de *kappa*+1. Luego se declara el vector *unsummed* como un vector de tamaño 4x1 en donde el primer elemento es cero y los otros elementos son la diferencia de los inputs *theta* y el vector *kappa*. El vector *probs* se define como un K-simplex[1] con la función $softmax$. Finalmente, dado el vector de probabilidades *probs*, con la función *categorical_lpmf* se genera el logaritmo de la función másica de probabilidad.

- **Bloque de programa: Datos**
  En este segundo bloque se declaran las variables que son interpretadas en la lectura de datos. Se declara el tipo de dato (real, entero, vector, matriz) y las restricciones en el dominio con el comando $< lower = 1 >$,

```
data {
int<lower=1> I;                    // # Items
int<lower=1> J;                    // # Personas
int<lower=1> N;                    // # Respuestas
int<lower=1,upper=I> ii[N];        // i for n
int<lower=1,upper=J> jj[N];        // j for n
int<lower=0> y[N];                 // Respuestas for n;
y = 0,1...m_i
}
```

  En este caso I corresponde al número de Ítems (15); J al número de personas (144), N al número total de respuestas (2160), los elementos ii y jj indican cada respuesta $y[n]$ a que ítem y persona pertenece, con dominio de valor máximo de I y J respectivamente.

- **Bloque de programa: Datos transformados**

  El objetivo de este bloque es declarar y definir variables que no necesitan ser cambiadas cuando se ejecuta el programa. Variables declaradas en el bloque de datos podrían ser usadas para declarar variables transformadas. En este caso se declara la variable entero m que corresponde el número de parámetros por ítem tomando la categoría máxima de los ítem.

```
transformed data {
int m;
m = max(y);}
```

- **Bloque de programa: Parámetros**

  Las parámetros declarados en esta sección serán las variables que Stan muestreará con sus muestreadores (HMC y NUTS). A las variables que se declaran como parámetros no se les puede asignar un valor directamente y al igual que en el bloque de los datos, acá se pueden plantear restricciones como por ejemplo un valor mínimo de cero $< lower = 0 >$ para el hiperparámetro de varianza $\sigma_{\kappa_{k,l}}$, el código en este trabajo para los parámetros corresponde a:

[1]Un K-simplex es definido por $softmax(y) = exp(y)/\sum_{k=1}^{K} exp(y_k)$

```
parameters {
vector<lower=0>[I] alpha;
ordered[m] kappa[I];
vector[J] theta;
real<lower=0> sig_the;
real mu_the;
}
```

En este código se expresan restricciones en un dominio positivo para el parámetro alpha de discriminación y para el hiperparámetro sigma $\sigma_{\kappa_{k,l}}$. El parámetro $\kappa_{k,l}$ se declara como un vector ordenado con el comando *ordered*[m] indicando que se tienen m puntos de corte que para esto caso es $m = 3$. El vector de parámetros de habilidad θ se declara como un vector con J componentes que corresponde al número de individuos y el hiperparámetro $\sigma_{\kappa_{k,l}}$ y $\mu_{\kappa_{k,l}}$ como valores reales con el comando *real*.

- **Bloque de programa: Modelo**

Como su nombre lo indica, acá se define el modelo de probabilidad y las distribuciones a priori para los parámetros, por eso en este paso es calculada la log densidad posterior. En este bloque se pueden definir variables seguido por sentencias, pero estas serán locales y no se presentaran en la salida del programa. Como se definió en (4-1), las distribuciones a priori son asignadas para los parámetros e hierparámetros.

```
model {
alpha ~ normal(0, 2);
for(i in 1:I)
kappa[i] ~ normal(mu_the,sig_the);
theta ~ normal(0,1);
mu_the ~ normal(0,1);
sig_the ~ cauchy(0,2);
for (n in 1:N)
target += pcm(y[n], theta[jj[n]].*alpha[ii[n]],
kappa[ii[n]]);
}
```

Es importante tener en cuenta que la parametrización presentada en este código para $\kappa_{k,l}$ en realidad representa $\alpha_k * \kappa_{k,l}$ planetado en la ecuación del modelo (4-1), por esta razón, para encontrar el verdadero valor para el parámetro de dificultad $\kappa_{k,l}$, se debe dividir en la salida de Stan el valor de $\kappa_{k,l}$ entre los parámetros de discriminación α_k. La función *pcm* definida en el bloque de funciones genera un vector de log probabilidades predichas por el modelo para las respuestas $Y[n]$. La verosimilud incrementa respectivamente con las log posteriores con el comando $target+=$.

- **Bloque de programa: Cantidades generadas**

  Este bloque es más bien diferente a los otros bloques ya que las declaraciones en este no afectan a los valores de los parámetros muestreados. Este módulo sólo se ejecuta cuando una muestra ha sido generada y dentro de las aplicaciones de la inferencia posterior que puede ser realizadas son: muestreo directo para generar datos y diagnosticar modelos, generar predicciones para nuevos datos y calcular probabilidades de eventos a posteriori entre otras.

```
generated quantities {
        ordered[m] beta[I];
        real pval_model;
        vector[N] Ty;              // Discrepance (data)
        vector[N] Ty_rep;          // Discrepance (rep)
        int<lower=0> y_rep[N];     // replica


        for(i in 1:I){
        for (n in 1:N){

        y_rep[n]   =   categorical_rng(pcm_1(y[n], theta[jj[n]].
        *alpha[ii[n]], kappa[ii[n]]));
        Ty[n]      =   pcm(y[n], theta[jj[n]].*alpha[ii[n]],
        kappa[ii[n]]);
        Ty_rep[n] =   pcm(y_rep[n], theta[jj[n]].*alpha[ii[n]],
        kappa[ii[n]]);}}
        pval_model = step(sum(Ty)-sum(Ty_rep));
        }
```

En este trabajo se empleó el modulo de cantidades para calcular el p-valor bayesiano generando datos de las muestras de los parámetros de la posterior para luego calcular un factor de discrepancia.

B. Anexo: Constructo Inicial

El instrumento inicial como se explicó en la sección cinco, estaba compuesto por 18 ítems polítomicos ordenados (A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R) que evalúan diferentes características económicas, comportamentales y de conocimiento financiero del encuestado. El objetivo de dichos ítems es construir una medida asociada al perfil de riesgo de los inversionistas, con el fin de formular y perfilar estrategias de inversión adecuadas para los agentes. A continuación se presentan los ítem originales aplicados en la encuesta:

(A) Al momento de invertir, usted prefiere productos financieros que devuelvan los retornos generados en un plazo de:

  (i) Menos de 1 año

  (ii) Entre 1 año y 3 años

  (iii) Más de 3 años

(B) ¿Con cuál de estas etapas financieras se identifica mejor actualmente?

  (i) No tengo un patrimonio, solvento mis gastos con mis ingresos mensuales

  (ii) Estoy ahorrando y empezando a construir mi patrimonio

  (iii) Ya tengo un patrimonio y lo estoy fortaleciendo

(C) Asumiendo que va a realizar una inversión ¿Qué porcentaje de su patrimonio actual destinaría para esto?

  (i) Más del 40 %

  (ii) Entre el 20 % y el 40 %

  (iii) Menos del 20 %

(D) Teniendo en cuenta que en el mercado financiero una rentabilidad alta implica un riesgo alto, preferiría productos con una:

  (i) Rentabilidad baja

  (ii) Rentabilidad moderada

  (iii) Rentabilidad alta

(E) Frente a posibles desvalorizaciones en los mercados financieros que puedan afectar el valor de su inversión, usted preferiría:

(i) Retirar totalmente la inversión para evitar perdidas

(ii) Mantener la inversión fijando un límite de pérdida, ya que posiblemente se recupere el mercado

(iii) Incrementar la inversión, ya que en ese momento los precios son bajos, y si el mercado se recupera las ganancias serían mayores

(F) Usted ha comprado acciones de cierta empresa y un mes después el precio de esas acciones suben un 25 %. ¿cómo procedería ante esta situación?

(i) Venderlas ahora mismo para materializar las ganancias

(ii) Mantenerme y acumular dichas ganancias

(iii) Comprar más acciones, ya que si el precio sigue aumentando las ganancias serían elevadas

(G) Acaba de ganar un premio en un casino. ¿Qué opción elegiría para reclamarlo?

(i) $ 1.000.000 COP en efectivo

(ii) Se lanza una moneda, si cae cara reclama $6.000.000 COP si cae sello reclama 0$ COP

(iii) Se lanza un dado, si cae la cara con el número 2 ó el número 5 reclama $10.000.000 COP si cae otra cara reclama 0$ COP.

(H) Su mejor amigo le cuenta sobre una buena oportunidad de inversión en la Bolsa de Valores, pero usted tiene que pedir prestado el dinero para poder invertir y tiene poco tiempo para tomar la decisión. ¿Tomaría usted un préstamo?

(i) Definitivamente no

(ii) Tal vez, pero un monto no tan elevado

(iii) Sí claro, confío el concepto de mi amigo

(I) ¿En qué rango de edad se encuentra?

(i) Más de 40 años

(ii) Entre 30 y 40 años

(iii) Menos de 30 años

(J) ¿Cómo describe su expectativa de ahorro para los próximos 5 años?

(i) No se elevará ya que tengo inversiones y gastos planeados

(ii) Será estable ya que mis gastos e ingresos se mantendrán

(iii) Es probable que aumente mi capacidad de ahorro

(K) De sus ingresos mensuales, su capacidad de ahorro estimada es:
   (I) Menos del 20 %
   (II) Entre el 20 % y 40 %
   (III) Más del 40 %

(L) Su experiencia o conocimiento con productos CDT's es:
   (I) Ninguna, prefiero asesoría.
   (II) Básica y prefiero asesoría.
   (III) Suficiente para realizar una inversión.

(M) Su experiencia o conocimiento con productos relacionados con bonos (títulos emitidos por entidades privadas o públicas) es:
   (I) Ninguna, prefiero asesoría.
   (II) Básica y prefiero asesoría.
   (III) Suficiente para realizar una inversión.

(N) Su experiencia o conocimiento con productos relacionados con acciones es:
   (I) Ninguna, prefiero asesoría.
   (II) Básica y prefiero asesoría.
   (III) Suficiente para realizar una inversión.

(O) Su experiencia o conocimiento con productos relacionados con fondos de inversión colectiva o fondos de pensión es:
   (I) Ninguna, prefiero asesoría.
   (II) Básica y prefiero asesoría.
   (III) Suficiente para realizar una inversión.

(P) Su experiencia o conocimiento con productos derivados, divisas o activos de renta variable
   (I) Ninguna, prefiero asesoría.
   (II) Básica y prefiero asesoría.
   (III) Suficiente para realizar una inversión.

(Q) ¿Cómo clasificaría su experiencia en inversiones?
   (I) Limitada, tengo poca experiencia en inversiones
   (II) Moderada, tengo alguna experiencia, pero me gustaría recibir orientación y asesoría en la medida que lo solicite

   (iii) Suficiente, me siento seguro para tomar decisiones de inversión y estoy facultado para entender y ponderar los riesgos asociados.

(R) ¿Ha tomado capacitación o algún tipo de orientación respecto a las posibilidades de inversión en la banca o en el mercado de valores colombiano?

   (i) No, nunca

   (ii) Sí, pero de forma muy superficial

   (iii) Sí, me he asesorado y conozco las principales posibilidades

Bibliografía

[Bock and Lieberman, 1970] Bock, R. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35:179–197.

[Brentari and Golia, 2007] Brentari, E. and Golia, S. (2007). Unidimensionality in the rasch model: how to detect and interpret. *Statistica*, 67:253–261.

[Caviezel et al., 2011] Caviezel, V., Bertoli, L., and Lozza, S. (2011). Measuring risk profile with a multidimensional rasch analysis. *Journal of Applied Quantitative Methods*, 6:14–29.

[D, 1943] D, L. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, Series A*, 23:273–287.

[Edelen and Reeve, 2007] Edelen, M. and Reeve, B. (2007). Applying item response theory (irt) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*, 16:5.

[Fox, 2010] Fox, J. P. (2010). *Bayesian Item Response Modeling*. Springer, New York.

[Gao, 2018] Gao, G. (2018). *Bayesian Claims Reserving Methods in Non-life Insurance with Stan*. Springer.

[Gilliam et al., 2010] Gilliam, J., Chatterjee, S., and Grable, J. (2010). Measuring the perception of financial risk tolerance: A tale of two measures. *Journal of Financial Counseling and Planning*, 21.

[Grabble and Lytton, 1999] Grabble, J. and Lytton, R. (1999). Financial risk tolerance revisited: The development of a risk assesment instrument. *Financial Services Review*, 8:163–181.

[Grabble and Schumm, 2010] Grabble, J. and Schumm, W. (2010). An estimate of the reliability of the survey of consumer finances risk-tolerance question. *Journa of Personal Finance*, 9:117–31.

[Hambleton et al., 1991] Hambleton, R., Swaminathan, H., and Rogers, H. (1991). *Fundamentals of Item Response Theory*. Sage, London.

[Hoffman and Gelman, 2014] Hoffman, M. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623.

[Janseen et al., 2014] Janseen, G., Meirer, M., and J, T. (2014). Classical test theory and item response theory:two understandings of one high-stakes performance exam. *Colombian Applied Linguistics Journal*, 16:167–184.

[Linacre J, 1994] Linacre J, M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4):328.

[Lord, 1952] Lord, F. (1952). A theory of test scores. *Psychometric Monograph*, (No 7).

[Masters, 1982] Masters, G, M. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47:149–174.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.

[Monila, 2016] Monila, J. (2016). Análisis de distribuciones a priori de los parámetros de escala del modelo de regresión poisson inflado con ceros. Master's thesis, Universidad Nacional de Colombia.

[Monnahan et al., 2017] Monnahan, C., Thorson, J., and Branch, T. (2017). Faster estimation of bayesian models in ecology using hamiltonian monte carlo. *Methods in Ecology and Evolution*, 8:339–348.

[Muraki, 1992] Muraki, E. (1992). A generalized partial credit model: Aplication of the em algorithm. *Applied Psychological Measurement*, 16(2):159–176.

[Ostini and Nering, 2006] Ostini, R. and Nering, M. (2006). *Polytomus item response theory models*. Sage pub, London.

[Rangel, 2019] Rangel, H. (2019). Modelo politómico unidimensional de teoría de respuesta al ítem con distribución asimétrica de trazo latentene. Master's thesis, Universidad Nacional de Colombia.

[Rasch, 1960] Rasch, G. (1960). *Probabilistic Models for some Intelligence Tests and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.

[Roszkowski, 1992] Roszkowski, M. (1992). How to asses a client's financial risk tolerance: The basics personal financial risk tolerance. Master's thesis, Bryn Mawr, PA: The American College.

[Ruiz, 2016] Ruiz, M. (2016). *Usin Rasch Measurement Theory To Evaluate The Psychometric Quality Of a Financial Risk Tolerance Scale*. Tesis de doctorado, Universidad de Georgia, Athens.

[Samemija, 1997] Samemija, F. (1997). *The graded response model In W. J. van der Linden and R. K. Hambleton (Eds.)*. Handbook of Modern Item Response Theory, New York:Springer.

[Sherman et al., 2001] Sherman, H., Gutter, M., and Fan, J. (2001). A mesaure of risk tolerance based on economic theory. *Journal of Financial Counseling and Planning*, 12.

[Spearman, 1994] Spearman, C. (1994). The proof and measurement of assciation between two things. *American Journal of Psychology*, 15:72–101.

[Steve Brooks, 2012] Steve Brooks, Andrew Gelman, G. J. . X.-L. M. (2012). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC Press.

[Swaminathan and Gifford, 1982] Swaminathan, H. and Gifford, J. (1982). Bayesian estimation in the rasch model. *Journal of Educational Statistics*, 7:175–192.

[Wiberg, 2004] Wiberg, M. (2004). Classical test theory vs item response theory: An evaluation of the theory test in the swedish driving-license test. *EM: Educacional Measurement*, 50.